

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 11, 2002, 15:19:20 ; Search time 15 Seconds
(without alignments)
2514.189 Million cell updates/sec

Title: US-09-744-125-2

Perfect score: 5225

Sequence: 1 MALRGVRYKVLCAEKNDAK.....RKSLCHQPHGHRPCPNR 974

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5225	100.0	1001	1	TP3A_HUMAN
2	4586	87.8	1003	1	TP3A_MOUSE
3	2230.5	42.7	1250	1	TP3A_DROME
4	1711	32.7	759	1	TP3A_CAEEL
5	1279	24.5	622	1	TP3A_SCHPO
6	1174.5	22.5	656	1	TP3A_YEAST
7	1078	20.6	862	1	TP3B_MOUSE
8	1066	20.4	862	1	TP3B_HUMAN
9	1022	19.6	875	1	TP3B_DROME
10	688	13.2	663	1	TP3A_ARCFU
11	653	12.5	770	1	TP3A_THRAC
12	600.5	11.5	685	1	TP3A_PYRAB
13	591	11.3	686	1	TP3A_PYRHO
14	573	11.0	761	1	TP3A_METJA
15	563	10.8	718	1	TP3A_METHA
16	483.5	9.3	710	1	TP3A_LACIA
17	467	8.9	673	1	TP3A_AERPE
18	449.5	8.6	700	1	TP3A_CAMJE
19	441	8.4	690	1	TP3A_BACHD
20	441	8.4	876	1	TP3A_VIBCH
21	426	8.2	691	1	TP3A_BACSU
22	420	8.0	865	1	TP3A_SALTY
23	413	7.9	865	1	TP3A_ECOLI
24	412	7.9	1217	1	TP3A_ZYMO
25	406	7.8	898	1	TP3A_SYNY3
26	404	7.7	868	1	TP3A_HAETN
27	403.5	7.7	633	1	TP3A_THRMA
28	398	7.6	863	1	TP3A_PSEAE
29	396.5	7.6	868	1	TP3A_PASMU
30	382	7.3	1060	1	TP3A_PYRPU
31	379.5	7.3	776	1	TP3A_RICPR
32	377.5	7.2	736	1	TP3A_HELPY
33	375.5	7.2	776	1	TP3A_RICCN

ALIGNMENTS

34	370.5	7.1	952	1	TP3A_STRCO	09x909 streptomyc
35	370	7.1	660	1	TP3A_XYLFA	09phk2 xyliella fas
36	367	7.0	731	1	TP3A_TREPA	083409 treponema p
37	359.5	6.9	736	1	TP3A_HELPJ	09zmv7 helicobacte
38	356.5	6.8	861	1	TP3A_BUCAI	057371 buchnera ap
39	349	6.7	540	1	TP3A_AOUAE	066893 aquilex aco
40	344.5	6.6	650	1	TP3A_PASMU	09cp53 pasteurella
41	339	6.5	653	1	TP3A_ECOLI	p14294 escherichia
42	336.5	6.4	711	1	TP3A_MYCPN	p78032 mycoplasma
43	329	6.3	649	1	TP3A_SALTY	p40687 salmoneilla
44	328	6.3	647	1	TP3A_VIBCH	09kgf5 vibrio chol
45	325.5	6.2	651	1	TP3A_HAETN	p43704 haemophilus

RESULT 1

TP3A_HUMAN	STANDARD:	PRT:	1001 AA.
ID	Q13472: Q13473:		
AC	Q13472: Q13473:		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	DNA topoisomerase III alpha (EC 5.99.1.2).		
GN	TP3A OR TOP3.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96195027; PubMed=8622991;		
RA	Hanai R., Caron P.R., Wang J.C.;		
RT	"Human TOP3: a single-copy gene encoding DNA topoisomerase III.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 93:3653-3657(1996).		
CC	-1- FUNCTION: REDUCES THE NUMBER OF SUPERCOILS IN A HIGHLY NEGATIVELY SUPERCOILED DNA.		
CC	-1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded DNA, followed by passage and rejoining.		
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	-1- TISSUE SPECIFICITY: HIGH EXPRESSION IS FOUND IN TESTIS, HEART, SKELETAL MUSCLE AND PANCREAS.		
CC	-1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).		
CC	-----		
DR	EMBL: U43431; AAB03694.1; -		
DR	EMBL: U43431; AAB03695.1; -		
DR	EMBL: 601243; -		
DR	InterPro: IPR003601; DNATOP1_ATP_bind.		
DR	InterPro: IPR003602; DNATOP1_DNA_bind.		
DR	InterPro: IPR000380; Pro_topoisomerase.		
DR	InterPro: IPR002936; Toprim.		
DR	InterPro: IPR001878; Znf_CCHC.		
DR	Pfam: PF01131; Topoisom_bac; 1.		
DR	Pfam: PF01751; Toprim; 1.		
DR	Pfam: PF01396; zf-C4_Topoisom; 1.		
DR	Pfam: PF00098; zf-CCHC; 1.		
DR	PRINTS: PR00939; C2HCZNFINGER.		
DR	PRINTS: PR00417; PRTPISMRASET.		
DR	SMART: SM00437; TOP1ac; 1.		
DR	SMART: SM00436; TOP1bc; 1.		
DR	SMART: SM00493; TOPPRIM; 1.		
DR	SMART: SM00343; Znf_C2HC; 1.		

DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
 KM Isomerase; Topoisomerase; DNA-binding; Repeat; Zinc-finger;
 FT ACT_SITE 362 362 Polymorphism.
 FT DOMAIN 658 685 DNA CLEAVAGE (BY SIMILARITY).
 FT REPEAT 812 923 C4-TYPE (POTENTIAL).
 FT REPEAT 812 839 2 x 27 AA APPROXIMATE REPEATS.
 FT REPEAT 896 923 1.
 FT VARSPLIC 1 25 MISSING (IN SHORT ISOFORM).
 FT VARIANT 596 596 C->Y.
 FT SEQUENCE 1001 AA; 112372 MW; 06558C749569ECC2 CRC64;

Query Match 100.0%; Score 5225; DB 1; Length 1001;
 Best Local Similarity 100.0%; Pired. No. 0;
 Matches 974; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALGVRVLCVARKNDKAGIADLLSNGRRRRREGSKFKIYEPDVIKGNQNTVMT 60
 DB 28 MALGVRVLCVARKNDKAGIADLLSNGRRRRREGSKFKIYEPDVIKGNQNTVMT 87
 QY 61 SVSGHLAHDFQOMQFRKQSCNPLVFEAEIEKYCPENFVDIKTLERETQCALVTMT 120
 DB 88 SVSGHLAHDFQOMQFRKQSCNPLVFEAEIEKYCPENFVDIKTLERETQCALVTMT 147
 QY 121 DCDREGENIGETIIVCAVAPNLOVLRARSETPHAAVTRACENLTPEDORSADVVR 180
 DB 148 DCDREGENIGETIIVCAVAPNLOVLRARSETPHAAVTRACENLTPEDORSADVVR 207
 QY 181 QELRLRGAATRFQOTLRIQRPPEVLAEOILISGSCQPLTGFVVEFKIAQAVPELF 240
 DB 208 QELRLRGAATRFQOTLRIQRPPEVLAEOILISGSCQPLTGFVVEFKIAQAVPELF 267
 QY 241 HRIKVTDHKDGIVFEMKRRRLFNHRTACLVLYOLCVDPMATVVEVSKRSKRPQAL 300
 DB 268 HRIKVTDHKDGIVFEMKRRRLFNHRTACLVLYOLCVDPMATVVEVSKRSKRPQAL 327
 QY 301 DTVELEKIAKKLINAKEVRIAEKLYTQYISTPRTETNIFPRDLNLTVLVEQOTDP 360
 DB 328 DTVELEKIAKKLINAKEVRIAEKLYTQYISTPRTETNIFPRDLNLTVLVEQOTDP 387
 QY 361 RWGFAOSILIERGGPTPRNGKSDOAHPIHPRTYTNLQDEORLYEFIRHFLACSSQ 420
 DB 388 RWGFAOSILIERGGPTPRNGKSDOAHPIHPRTYTNLQDEORLYEFIRHFLACSSQ 447
 QY 421 DAQOQETVEIDIAQERFVAHGLMILARNYLDVYRYDHMSDKILPVYEGSGHFPQSTYEM 480
 DB 448 DAQOQETVEIDIAQERFVAHGLMILARNYLDVYRYDHMSDKILPVYEGSGHFPQSTYEM 507
 QY 481 VDGETSPKKLTEDDLIAEMKHGIGTDATAEHETIKAMNYGGLTPDKRFLPGHLGNG 540
 DB 508 VDGETSPKKLTEDDLIAEMKHGIGTDATAEHETIKAMNYGGLTPDKRFLPGHLGNG 567
 QY 541 LVEGYDSMGYEMKSPDLRAELADLKLCDGKKDFVLLROVOYKQVFLFEAAVAKAKL 600
 DB 568 LVEGYDSMGYEMKSPDLRAELADLKLCDGKKDFVLLROVOYKQVFLFEAAVAKAKL 627
 QY 601 DEALAQYFGNGTELAQOEDIYPAPEPIRKCPOCKNDKAVLTKKNGGFLVSCMGPEECRS 660
 DB 628 DEALAQYFGNGTELAQOEDIYPAPEPIRKCPOCKNDKAVLTKKNGGFLVSCMGPEECRS 687
 QY 661 AVLPLDSVLEASRDSVVCQPHVYRLKIKFRGSLPPTMPLEFYVCCIGGCDTLREI 720
 DB 688 AVLPLDSVLEASRDSVVCQPHVYRLKIKFRGSLPPTMPLEFYVCCIGGCDTLREI 747
 QY 721 LDRFSGGPPRASPQSRILANOSILNRMDNSHOPADSRQTSKALAQTLPPPTAGE 780
 DB 748 LDRFSGGPPRASPQSRILANOSILNRMDNSHOPADSRQTSKALAQTLPPPTAGE 807
 QY 781 SNSVTCMGQEAVALIYRKEGPNRGROFFKCGNGSCNFFLWADSPNPGAGPPALAYRPL 840
 DB 808 SNSVTCMGQEAVALIYRKEGPNRGROFFKCGNGSCNFFLWADSPNPGAGPPALAYRPL 867

QY 841 GASLCGCPGPIHLGCGFNPBGDSSGSGTSLCSQSPSVTRTYOKDGPNNKGRQFHTCAKPRE 900
 DB 868 GASLCGCPGPIHLGCGFNPBGDSSGSGTSLCSQSPSVTRTYOKDGPNNKGRQFHTCAKPRE 927
 QY 901 QOCGFQWDEMTAGTSGAPSWTDGRGTLESEARSRPRASSSDMGSTAKPKKCSIC 960
 DB 928 QOCGFQWDEMTAGTSGAPSWTDGRGTLESEARSRPRASSSDMGSTAKPKKCSIC 987
 QY 961 HOPGHTRPFCQNR 974
 DB 988 HOPGHTRPFCQNR 1001

RESULT 2
 TP3A_MOUSE
 ID TP3A_MOUSE STANDARD; PRT; 1003 AA.
 AC 070157;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DNA topoisomerase III alpha (EC 5.99.1.2).
 GN TOP3A OR Top3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA1B/C; TISSUE=Testis;
 RX MEDLINE=98201702; PubMed=9540825;
 RA Seki T., Seki M., Katada T., Enomoto T.;
 RT "Isolation of a cDNA encoding mouse DNA topoisomerase III which is
 RL Biochim. Biophys. Acta 1396:127-131(1998).
 CC -1- FUNCTION: REDUCES THE NUMBER OF SUPERCOILS IN A HIGHLY NEGATIVELY
 CC SUPERCOILED DNA.
 CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
 CC DNA, followed by passage and rejoining.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS.
 CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB006074; BAA25662.1; -;
 DR MGD; MGI:1197527; Top3a.
 DR InterPro; IPR003601; DNAtopI_ATP_bind.
 DR InterPro; IPR003602; DNAtopI_DNA_bind.
 DR InterPro; IPR003603; Pro_topoisomerase.
 DR InterPro; IPR002936; Toprim.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF01131; Topoisom_bac; 1.
 DR Pfam; PF01751; Toprim; 1.
 DR Pfam; PF01396; zf-C4_Topoisom; 1.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR PRINTS; PRO0039; C2HCZNFINGER.
 DR PRINTS; PRO0417; PRPISMRASEI.
 DR SMART; SM00437; TOP1AC; 1.
 DR SMART; SM00436; TOP1BC; 1.
 DR SMART; SM00493; TOPRIM; 1.
 DR SMART; SM00343; ZNF_C2HC; 1.
 DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
 KM Isomerase; Topoisomerase; DNA-binding; Repeat; Zinc-finger.
 FT ACT_SITE 362 362 DNA CLEAVAGE (BY SIMILARITY).
 FT DOMAIN 658 685 C4-TYPE (POTENTIAL).
 FT REPEAT 814 925 2 x 27 AA APPROXIMATE REPEATS.
 FT REPEAT 814 841 1.

```

FT REPEAT      898      925      2.
SQ SEQUENCE    1003 AA: 112358 MM: 254C738E746DE495 CRC64:

Query Match      87.8%; Score 4586; DB 1; Length 1003;
Best Local Similarity 87.4%; Pred. No. 8.8e-289;
Matches 853; Conservative 46; Mismatches 75; Indels 2; Gaps 1;

OY 1 MALRGYKVLCAVEKNDAAKGIADLLSNGMRREGLSKFNKIYEPFHYHLYGQNVNTVMT 60
DB 28 VARGRKVLCAVEKNDAAKGIADLLSNGMRREGLSKFNKIYEPFHYHLYGQNVNTVMT 87
OY 61 SVSGHLLAHDFQMFQFRKWSQCNPLVLEAEIEKYCPENFVADIKKTLERETROCCALYWT 120
DB 88 SVSGHLLAHDFQMFQFRKWSQCNPLVLEAEIEKYCPENFVADIKKTLERETROCCALYWT 147
OY 121 DCRBGENIGFEIITHCAKAVKPLVYLARAFSEITTPHVAFTACENLTEDPQRRSDAVDR 180
DB 148 DCRBGENIGFEIITHCAKAVKPLVYLARAFSEITTPHVAFTACENLTEDPQRRSDAVDR 207
OY 181 QEDLRIGAFTPRPOTLRQIRFEVLAEDLISGSCOPTLIGFVVERFAIAQFAVEIF 240
DB 208 QEDLRIGAFTPRPOTLRQIRFEVLAEDLISGSCOPTLIGFVVERFAIAQFAVEIF 267
OY 241 HRIKVTDHDKGIYEFNMKSHRLFNHTACLVLYOLCYEDPMATVEVRSKPKSKMRPQAL 300
DB 268 HRIKVTDHDKGIYEFNMKSHRLFNHTACLVLYOLCYEDPMATVEVRSKPKSKMRPQAL 327
OY 301 DYVELEKLASRKLRINKAKETMRIAEKLYTGYISYPTETNIPRDINTLVLYEQOTPD 360
DB 328 DYVELEKLASRKLRINKAKETMRIAEKLYTGYISYPTETNIPRDINTLVLYEQOTPD 387
OY 361 RMKAFPOSTLERGGPPRNKNSQOAHPRPHPRKYTNNGDDEORLKEFIVRHFLACCSQ 420
DB 388 RMKAFPOSTLERGGPPRNKNSQOAHPRPHPRKYTNNGDDEORLKEFIVRHFLACCSQ 447
OY 421 DAQOETTFEIDIAOERFVAHGLMIILARNYLDVYPYDWSDKILPYVEQSGSHFQSTVEM 480
DB 448 DAQOETTFEIDIAOERFVAHGLMIILARNYLDVYPYDWSDKILPYVEQSGSHFQSTVEM 507
OY 481 VDEGTSPKLLTEADLIALMEKHGIGTDATHAEIETIKARMYGLTTPDKRFLRGHLMG 540
DB 508 VDEGTSPKLLTEADLIALMEKHGIGTDATHAEIETIKARMYGLTTPDKRFLRGHLMG 567
OY 541 LVEGYDSMGYEMSKPDRAELDLKLCIGCKDKFEVLRQOYQKYTOVFLEAVAKAKL 600
DB 568 LVEGYDSMGYEMSKPDRAELDLKLCIGCKDKFEVLRQOYQKYTOVFLEAVAKAKL 627
OY 601 DEALAQYFGNGTELAQOEDIPYAMPEDIRKCPQCNKDMVLTKKNGSFYLSGMEFPCRS 660
DB 628 DEALAQYFGNGTELAQOEDIPYAMPEDIRKCPQCNKDMVLTKKNGSFYLSGMEFPCRS 687
OY 661 AWWLPDSVLEASRDSVCPYCOHPYVRLKIKRGSLLPMPLEFYCCGCCDDTLREI 720
DB 688 AWWLPDSVLEASRDSVCPYCOHPYVRLKIKRGSLLPMPLEFYCCGCCDDTLREI 747
OY 721 LDIFSGGPPRASPGSGLANOSLNRMDNSOH--POPADSRQSGKALAQTLPPPTAA 778
DB 748 LGLRFPALPRASOPSHLOLSQALNMDSSQHLNLSOPLVNRKTRPSKYTAQALLPPTTA 807
OY 779 GESNSVYTCNGOEAVLLTVRKEGPNRGRQFCKNGSGCNFLMADSPNPGAGPALAYR 838
DB 808 GESNSVYTCNGOEAVLLTVRKEGPNRGRQFCKNGSGCNFLMADSPNPGAGPALAYR 867
OY 839 PLGASLACPPGPGIHLGFGNPGDSSGSGISCLQSPSVTRTVOKDGPNGKROFHTCAK 898
DB 868 PPSVSVCCPSVSVSHMGFSLGSDSGTGPCGCPAVVTRTVOKDGPNGKROFHTCAK 927
OY 899 REQCCGFFQWVDENTAGTSGAPWTGDRGTLSEARSKRPRASSDSMGSTAKPKRCS 958
DB 928 REQCCGFFQWVDENTAGTSGAPWTGDRGTLSEARSKRPRASSDSMGSTAKPKRCS 987
OY 959 LCHQPGHTRFPCPNR 974
DB 974 LCHQPGHTRFPCPNR 974

```

```

DB 988 LCHQPGHTRFPCPNR 1003

RESULT 3
ID TP3A_DROME STANDARD; PRT; 1250 AA.
AC G9NG98; 09VIV1;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN TOP3-ALPHA OR CG10123.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RP SEQUENCE FROM N.A.
RA Plank J.L., Reineke J.C., Wilson T.M., Hsieh T.-S.;
RT "Drosophila melanogaster topoisomerase III alpha.";
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-Y., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson R.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jattali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclid J.M.,
RA Palazotto M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weissstock G.M., Wu D., Yang S., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhu Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: WEAKLY RELAXES NEGATIVE SUPERCOILS AND DISPLAYS A
CC -1- DISTINCT PREFERENCE FOR BINDING SINGLE-STRANDED DNA.
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC -1- DNA, followed by passage and rejoining
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE

```

```

CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AF255733; AAF1288.1; ALT SEQ.
DR      EMBL; AE003663; AAF53813.1; AUT SEQ.
DR      FlyBase; FB000040268; Top3-alpha.
DR      InterPro; IPR003601; DNATopI_ATP_bind.
DR      InterPro; IPR003602; DNATopI_DNA_bind.
DR      InterPro; IPR000380; Pro_topoisomerase.
DR      InterPro; IPR002936; TopIrm.
DR      InterPro; IPR001878; znf_CCHC.
DR      Pfam; PF01131; Topoisom_bac; 1.
DR      Pfam; PF01751; TopIrm; 1.
DR      Pfam; PF01396; zf-C4_Topoisom; 1.
DR      PRINTS; PR00417; PRPISOMRASEI.
DR      SMART; SM00437; TOP1BC; 1.
DR      SMART; SM00436; TOP1BC; 1.
DR      SMART; SM00493; TOPRIM; 1.
DR      SMART; SM00343; znf_CZHC; 1.
DR      PROSITE; PS00396; TOPOISOMERASE_1_PROK; 1.
DR      Isomerase; Topoisomerase; DNA-binding.
KW      ACT SITE 356 356 DNA CLEAVAGE (BY SIMILARITY).
SQ      SEQUENCE 1250 AA; 136137 MW; 28809F70B3DB75E CRC64;

Query Match 42.7%; Score 2230.5; DB 1; Length 1250;
Best Local Similarity 41.6%; Pred. No. 2,1e-136;
Matches 514; Conservative 125; Mismatches 307; Indels 291; Gaps 30;

QY      8 KVLCAEKNDAKAGIADLSNGMRRRRGLSKFNKTYEEDYHLYGQONTMWTSSGHL 67
DQ      27 KYLNAEKNDAKATAGLISNGAORREGYSYNNKVPFEADVRQNMKNVTSVGHMM 86
QY      68 AHDFQMRKMOQSNPLVLEAEIKCYRPENFVDIKKTLERTRQQAOLVITWDCDREG 127
DQ      87 QLAQVSYKKNMRTVPDRSLFADPAVEKGVGSDEPKIKRTLERVRRCQGLIITWDCDREG 146
QY      128 NIGEETIHVCAKVDNLOVLRASESEITPHAVRTACENLTPEQDVSPADVROELDIRI 187
DQ      147 NIGVEIIVCAIKKFNISYVRATFSEITTVAVRRALQOLGQPKQSDPAVDVRELDLRT 206
QY      188 GAATFRQDTLRLQRFPEVLAEQSLSYSSCOPTLGFVVERPKALQAAVPELFIKTYH 247
DQ      207 GAATFRQDTMRLQRFPEKRIADKLISYSSCOPTLGFVAERKLEAVVSEPFMIKLYH 266
QY      248 DHKQDIEFNKRRHFLFNHTACLYVLCV--EDPMATVVEVRSKPKSKMRRQALDVEL 305
DQ      267 TIDDLTVEFNMAARNLFDKEACENLLICLAEPDRALVESYTVPKKMKRPTLDIYEM 326
QY      306 EKLSARKLRIAKETMRIAEKLYTGYISYPTETINIEPRDLNLTVLVEQOTPPDRMGAF 365
DQ      327 EKLGRSLKLSAKETMTIAEKLYTGFTSYPTETINQSKKEFLAPLVEMQGHDRMGAF 386
QY      366 AQSILIEGGRPPRNGKSDQAHPPHPKRYTNLOGDEORLYELFVIRFLACSSDAQOQ 425
DQ      387 AQRVIE-WGPNPDRNGKSDQAHPPHPKRYTNLOGDEORLYELFVIRFLACSSDAQO 445
QY      426 ETVYVIEDIAQEFVHGLMILARNLVLDVYPRVDHMSDKILPYVEQGSHPQSPVEVDEGT 485
DQ      446 ETVYVIEDIAQEFVHGLMILARNLVLDVYPRVDHMSDKILHINENQRRPEPEVLSLHGAT 505
QY      486 SPKKLTLEADLIALMEKKGIGTDATNHAHETIKARMVVGLTPDRKRF-PGHLGMLVEG 544
DQ      506 TAPRLTLEADLIALMEKKGIGTDATNHAHIMTKRGYIGVL-DKGLFVPGYIGMGLVEG 564
QY      545 YDSMGYEKSPDPLRALEADLKLCDGKKDKFVVLROQVQYKYQVETAVAKAKKLDEL 604

```

Db	565	YAMETALATAKPOLRAHEFLDLKLICGQKDPKRVVLTELDIAKKQAVQQTITDKITAMDKI	624
Qy	605	AOYFGN--GTEIAQOEDDIYPAMP-----EPIKCPQCNK-DMYETKTKN-GGFYLSMGE	655
Db	625	SARINETPAAANSAYOEGADGASAPISGIIIOISIFQCEPKCNDEAPLALPKRKNQOGWYIGCNMF	684
Qy	656	PCRSAVMPLPDSVLASNRSSVCPYQCPHPYRLKLTKFRKSLP-----PTMLEFPC	709
Db	685	PCKNAVMWLPTECKKASVLDCECPCPGDG--YRM-LKF-RLSTPYRGVFGTPSGWYKTC	740
Qy	710	IGCDTLEILDLRF-----SGGP-----	729
Db	741	L-PCDNLFTTTNINLDSYKKGIVGEYKRGGGGPGFGPGGGSGRAGSGGSSGPG	799
Qy	730	-----	729
Db	800	GGSGGGGSGGSGGTGGGGSGGSGTGCGGLGCGKPKPGESKSGATKPPNEPK	859
Qy	730	---FRA-----SQPSRLQA-----NQSINRND-----	749
Db	860	KTKERKAPPNKTKTSSKSGSIRSFPTSAAPTNSASNGIDEFDSNDGEDAMLAESVE	919
Qy	750	NSQHO-----PAD-----	758
Db	920	SSQPRITISMPDLDDIAAFADDAEAEALVNGCTPMTESNGDQQLDKSLSEIKQD	979
Qy	759	-----SKOTGSSKALQTLF-----PTPAGESNSVYC-NGC	789
Db	980	KADERPMLGTERASLSLGTAAPTPPKPAKRKMDSYERDSTPSSVSEPTVLCGTCQ	1039
Qy	790	QEAULLYRKESPEPNGRQFFKC-NGSCNFPLMADSPNPGAGGPALALYRLG-----ASL	844
Db	1040	QPARNTVYRKKNPNGLRLLYKCPKDECNFQWADDEPPSASKSKNSTGSAOSTTSMGSN	1099
Qy	845	GCPEPGIHLG-----GFGNPGDGSQSGTSCLSQSP	875
Db	1100	RYVTLPSIQSNSQSGSSMSBNSSTYTITQTKTKQDERNTAIPGD--GEYWCNCGQL	1157
Qy	876	SVRTVQKDPKPKQKQFTCAKPRBQCGCFQWVDENTAPGTSCAPSWTGDGRTLSEEA	935
Db	1158	ASQLVYRKDGPNQGRPYAC-PTREKSCGFPEKMGEDQNGAS-STSW-----GSANRMP	1211
Qy	936	RSKRPRASSDMGSTAKKPKKCSLCHQGHRRPFCQ	972
Db	1212	GRSOPTAITSQD---GPKTRRCGLCRKGHTRNKCPR	1244
RESULT 4			
TOP3_CAEEL	TOP3_CAEEL	STANDARD:	PRT: 759 AA.
AC	O61660;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	DNA topoisomerase III (EC 5.9.9.1.2).		
TOP3			
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;		
OC	Rhabditiidae; Peloderiinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL NZ;		
RA	Kim Y.-C., Koo H.-S.;		
RT	*cDNA cloning and overexpression of Caenorhabditis elegans DNA		
RT	topoisomerase III.*;		
RL	Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.		
CC	-I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded		
CC	DNA, followed by passage and rejoining.		
CC	-I- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE		
CC	FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright It is produced through a collaboration		

	Matches	253;	Conservative	127;	Mismatches	198;	Indels	22;	Gaps
Qy	8	KVLVAEAKNDAAKGIADILSLNGRRRRREGGLSKFNKITYEFYHLYGQV--	TMVTSVSG	64					
Db	2	RVLCAEAENKSIAKSVASIIIGGVHVRRRDRRTKYKKNIDFSFN -EGGVNSDDVTMTSVSG		60					
Qy	65	HLHMHDFQMRKMOSCPVLVFAEIEEKYCPENFVVDIKKLEBETROCALVITWTCDR		124					
Db	61	HLTEASFSEFSESSMSVSVQDVLFPQAQIITTSYKNAEVLADIKKEARNAOYLITWTCCDR		120					
Qy	125	EGENIGFEIHHYCAKVKPNLOYLPARFSEITPHAVKPAACENLTPEDDORSADVADVRQELD		184					
Db	121	EGEHIGVEISVVAASNSIQIVIRADNNLERSHIIISAKKRPDVSNNADAVADARFIELD		180					
Qy	185	LRIQAFRPTLRLORFRFPELVLEOLISVSGCOPTLGFVFERFKAIOAFVPEIHRIK		244					
Db	181	FRLGAIFFRLQITQLQNSF -DILONKIIISTGSPCPPLTGFVVDWRQVEDVPEPTWHLR		235					
Qy	245	VTHDHKDGIVFENMKRHRLEFNHTACLYLYOLCVSDPMATVVEVSKPKSRMRRQALDIVE		304					
Db	240	FVDRKQGTIOENMERAKVFRPLPTMIILENCLECKTAKAVNITQKPKTKKPLPLSTVE		299					
Qy	305	LEKLASRKLRIINAETMRMAIAKLYITQGYISIPREFTNIFPDLNLVLYLVEDQTPDPWGA		364					
Db	300	LTKRGPKRLRISAKTKLELAENLVTNGVSPVPRTEQDPOSSNNLHAIIDKLTGAQWSDS		355					
Qy	365	FAQISLEGSGTTPNGNKSDDAPHPRIPTKTYTNL---QGDEQLYEFYVHIFACSD		421					
Db	360	YAEGLLAGDYRPRPKRKHNDRAHRRPHVQGVVHNSALPQSDHKMKVYELLIRRLFACSSDN		419					
Qy	422	AGQGETTYEIDIAQERVAHSLMLIARNYLDVYYDIW -SPKILPVYEQGSHPOPSIVEM		480					
Db	420	AKGAGETIYQVMEELFESKCLLTETKNYLEVYVEKWMESDDQLEPYRLHEEPQPLHDM		479					
Qy	481	VDGETSPKLTLEADLALMEKHGIGDPAFAEHIEFIKAMVY-----GLTPDKRF		533					
Db	480	MDSSTSSSYITTEBELTALMDANIGIDATMAHIEIKVEYEREYVIRKKKKGQGYT---EF		538					
Qy	533	LPHGILMGVLEGYDVSQGYE -MSKPRPLRAELIADLKLICDGKKDKFYVLROQVQKYQVF		590					
Db	537	VPSISIGVALAKGYEIGLEWISLTPEFPRKKEMEVQLKNIENGQLNRNVLVHMLITQERDFV		596					
RESULT 6									
TOP3_YEAST	TOP3_YEAST	STANDARD;	PRT;	656	AA.				
AC	P13099;								
DT	01-JAN-1990 (Rel. 13, Created)								
DT	01-JAN-1990 (Rel. 13, Last sequence update)								
DT	01-OCT-1996 (Rel. 34, Last annotation update)								
DE	DNA topoisomerase IIi (EC 5.99.1.2).								
GN	TOP3 OR EDR1 OR YLR234W OR L6083.3.								
OS	Saccharomyces cerevisiae (Baker's yeast).								
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;								
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.								
OX	NCBI_Taxid=4932;								
RP	SEQUENCE FROM N.A.								
RC	MEDLINE=69324087; Pubmed=2546682;								
RA	STRAIN=S288C / AB972;								
RA	Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,								
RA	Favellon A., Fulton L., Gattung S., Greco T., Kirsten J.,								
RA	Kucaba T., Hallsworth K., Hakkis J., Hillier L., Jier M.,								
RA	Johnson D., Johnston L., Langston Y., Latreille P., Le T.,								
RA	Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,								
RA	Wardle L., Riles L., Taich A., Trevasakis E., Vignati D.,								
RA	Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;								

```

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 578-592, 605-615, 794-805 AND 1038-1046.
RA MEDLINE=93330283; PubMed=8336724;
RX Jiang W., Middleton K., Yoon H.-G., Fouquet C., Carbon J.;
RT "An essential yeast protein, Cbf5p, binds in vitro to centromeres and
RL microtubules.";
RL Mol. Cell. Biol. 13:4884-4893(1993).
CC
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER. ESSENTIAL
CC FOR PROPER CHROMOSOME SEGREGATION IN BOTH MEIOSIS AND MITOSIS.
CC WEAKLY RELATES NEGATIVE SUPERCOILS AND DISPLAYS A DISTINCT
CC PREFERENCE FOR BINDING SINGLE-STRANDED DNA. INTERACTS WITH SCS1.
CC THE TOP3-SCS1 PROTEIN COMPLEX MAY FUNCTION AS A EUKARYOTIC REVERSE
CC GYRASE INTRODUCING POSITIVE SUPERCOILS INTO EXTRACHROMOSOMAL
CC RIBOSOMAL DNA RINGS.
CC
CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC
CC -I- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC -----
CC EMBL; M24939; AAA35161.1; -;
CC EMBL; U19027; AAB67406.1; -;
CC PIR; A33169; ISBT3.
CC SGD; S0004224; TOP3.
CC InterPro; IPR003601; DNATOP1_ATP_bind.
CC InterPro; IPR003602; DNATOP1_DNA_bind.
CC InterPro; IPR000380; Pro.topoisomerase.
CC InterPro; IPR002936; Top1im.
CC Pfam; PF01131; Topoisom_dac; 1.
CC Pfam; PF01751; Top1im; 1.
CC PRINTS; PR00417; PRTPISMASE1.
CC SMART; SM00437; TOP1AC; 1.
CC SMART; SM00436; TOP1BC; 1.
CC SMART; SM00493; TOP1RM; 1.
CC PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
CC
CC KM Isomerase; Topoisomerase; DNA-binding.
CC ACT_SITE 356 356 DNA CLEAVAGE (BY SIMILARITY).
CC SO SEQUENCE 656 AA; 74370 MW; 51DF78936A8BBA4 CRC64;
CC
CC -----
Query Match 22.5%; Score 1174.5; DB 1; Length 656;
Best Local Similarity 41.8%; Pred. No. 1,66-66;
Matches 264; Conservative 104; Mismatches 213; Indels 51; Gaps 12;
QY 8 KVLCAEKNDAKGIADLLNSGRMRREGLSKFNKTYEFDYHL- - - - -GQNTVMVTSV 62
  |||||  : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 KVLCAEKNDAKGIADLLNSGRMRREGLSKFNKTYEFDYHL- - - - -GQNTVMVTSV 61
QY 63 SGHLLAHDFQMFQKMGQCNPLVLEAEIEKVCPENFYDIKKLEBRETROQQAIVIMTDC 122
  |||||  : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 AGHLLAGIDESHDSHGWCALIQLELDAPLAINMNNOKKIASNTKREARNADYIMTDC 121
QY 123 DREGNIGFETLIHWCK - - - - -AVKPNIOVLRFAFSEITPHAVTACENTLEPDORVSDADV 179
  |||||  : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 DREGYIGWETMEQKGRNRLIQNDQYRAYVSHLERQHLLMAARNPSRLDMKSVHAGVT 181
QY 180 KOELDLRIGAAFTPRQT - - - - -LRDRIPEVLAE - - - - -QLISYSGCPPLGF 224
  |||||  : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 RIEDLRAGYVETFLITETLRNKLARNQATMTKDAKHGCKNDQSVYSGTCQCPPLGF 241
QY 225 VVERFKALQAVPELPHRIKVTYHDKD - - - - -GTFENMKRHRLEFNACVLYQLVEDP - - 280
  |||||  : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 VDPRFERLRNVPPEFWYIQLVYENKQNGGTTTQMDRGHLEFDLSVLTFTFCIEETAGN 301
QY 281 MATVEVNSKRSKWRPDQDVTLEELKASRKLINAKETMRIAEKLYTGQYISYPTRET 340

```



```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN RP SEQUENCE FROM N.A.
RX RA MEDLINE=99128286; PubMed=9927731;
RT RT "Ng S.-W., Liu Y., Hasselblatt K.T., Mok S.C., Berkowitz R.S.;
RL RL "A new human topoisomerase III that interacts with SGS1 protein.";
RN RN Nucleic Acids Res. 27:993-1000(1999).
RP RP [2]
RX RA SEQUENCE FROM N.A.
RT RT MEDLINE=97228902; PubMed=9074928;
RL RL Kawasaki K., Minoshima S., Nakato E., Shibuya K., Shintani A.,
RN RN Schmeits J.L., Wang J., Shimizu N.;
RT RT "One-megabase sequence analysis of the human immunoglobulin lambda
RL RL gene locus.";
RN RN genome Res. 7:250-261(1997).
RP RP [3]
RX RA SEQUENCE FROM N.A.
RT RT Riou J.F., Goulaouic H., Grondard L.;
RL RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN RN [4]
RP RP SEQUENCE FROM N.A.
RX RC TISSUE=Skin;
RL RL Strausberg R.;
RN RN Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC CC -I- FUNCTION: POSSESS NEGATIVELY SUPERCOILED DNA RELAXING ACTIVITY.
CC CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC CC DNA, followed by passage and rejoining.
CC CC -I- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3: ARE
CC CC PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT C-TERMINAL
CC CC DOMAINS.
CC CC -I- TISSUE SPECIFICITY: THE TOP3B ISOFORMS HAVE DIFFERENT TISSUE
CC CC SPECIFICITIES. ISOFORM 1 IS FOUND IN TESTIS, HEART AND SKELETAL
CC CC MUSCLE. A 4 KB TRANSCRIPT, WHICH PROBABLY REPRESENTS ISOFORM 2, IS
CC CC FOUND IN THYMUS, KIDNEY AND PANCREAS.
CC CC -I- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC CC FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce
CC CC or send an email to license@sib-sib.ch).
CC CC -----
DR DR EMBL, AF053082; AAD15791.1; -
DR DR EMBL, AF017146; AAD01614.1; -
DR DR EMBL, AF125216; AAD29670.1; -
DR DR EMBL, BC002432; AAH02432.1; -
DR DR HSSP, P06612; IECL.
DR DR MIM, 603562; -
DR DR InterPro: IPR003601; DNATOP1_ATP_bind.
DR DR InterPro: IPR003602; DNATOP1_DNA_bind.
DR DR InterPro: IPR00380; Pro_topoisomerase.
DR DR InterPro: IPR002936; Toprim.
DR DR Pfam, PF01131; Topoisom_bac; 1.
DR DR Pfam, PF01751; Toprim; 1.
DR DR PRINTS; PR00417; PRTPISMRASE1.
DR DR SMART; SM00437; TOP1AC; 1.
DR DR SMART; SM00436; TOP1BC; 1.
DR DR SMART; SM00493; TOP1RM; 1.
DR DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW KW Isomerase, Topoisomerase; DNA-binding, Alternative splicing.
FT FT ACT_SITE 336 336
FT FT VASPLIC 703 730
FT FT VASPLIC 731 862
FT FT VASPLIC 703 707
FT FT VASPLIC 708 862
FT FT CONFLICT 365 365
FT FT CONFLICT 862 AA; 96661 MW; 75532827856CFE8F CRC64;
FO SEQUENCE

```

```

Query Match Similarity      20.4%: Score 1066; DB 1; Length 862;
Best Local Similarity      36.3%: Pred. No.2.3e-61;
Matches      280; Conservative      109; Mismatches      306; Indels      76; Gaps      19

QY      6 VRKVLVLAERKNPAKAGIADLLNSGMRARRREGSLKPFKKIYEFDNHLGQNTVMYTSVSGH      65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1 MNTVLMVAEKPRSLAQSIAITLISRGSLSSHKGLNGACSVHEHYGTFRAGQYRFRMTSYVCH      60

QY      66 LLAHDFOMQFRKWSQCNPLVLF-EADIEKYCPENFVDIKKTLERTRQCALIWTDCOR      124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      61 VMTLDFELGKYNKWDKVDPALEFSQAETKEKKAAPKLNMYKFLQVEGRGCDYIYLMDCDK      120

QY      125 EGENIGEELIHNCKAAMPYQ-----VLRARSEITPHAVRTACENLTPEDDQAVNSA      176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      121 EGENIEFEVL---DAVLPMYKNAHGGEKTVFRFRFSSTIDTDCNMMAACIGEDHNEALS      177

QY      177 VDVROELDIRIGAAFRFOTLRLQIRFFPVLALEQLISYSCOPPTLGFVVERFKAIQATV      236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      178 VVARQGLDIRIGCAFREFQTKYFGQKYGSD-LDSLSLSEFPQCPPTLGFCEVENHDKIQSK      236

QY      237 PELFFH--RIKYTHDHKDGIVERNMKRHLRPHNTACLVLYQLCEVDPMATVVERKSPKSK      294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      237 PETVYWLQAKKVNTD--KDRSLLDMDQVRFVREDRIAMFLNMKLEKQAOVEATSRKEKAK      295

QY      295 MRPQADLYELEKELASRKIRINAKETMRFAEKLYTQGYISYPTERTENIPRDLNTLVLE      354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      296 QRPRLALNTYEMLRVASSLSLGMGRQAHMOTAEKRLYDGYISTYPTETLTHPENFDLKGSLR      355

QY      355 QQTPEPRMGAFAOSILERRGGPTPRNGKSDQAHPRPHPTKYN--NLQSEQRLYEFIVR      412
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      356 QQANHPYMAADTVKRLLAEGINRPKRCHDAGD--HPITPMKSAATEALGGDAMRLVEYITR      414

QY      413 HFLACSQDAQOQETVELDIDQERFVAGMLANLYNDVYPIYDWS---DKILPYVE      468
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      415 HFLATVSHCKYLOSTISRIGPELTGSGKTVLSGFEYVMP---WQSVPLEESLPTQO      471

QY      469 QCSHPQPSVFEWVWDGTSPPKLTLEADLLALMEKHIGTDAFHAHIEITIKARMYGLTP      528
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      472 RQDAFPGVEKMLKEQOTNPDPYLTEAELLTMBKHIGTDAIPVAINNICQKRNVTYVBS      531

QY      529 DKRFLRGLMGLVBSGYDSMGYEMSKPDRLAELEADLKILCDGKKDFVYLROOYQYKQ      588
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      532 GRRLPTNIGILVLYGVYYIDAELVLPTRISAVERQOLNLAQGGKADYRQGLTGLDVEFR      591

QY      589 VRIEAVAKKIKDELAEQAFGNGTELAQEDDIYPRAMPREIRKCPQCNKMVLTKTKNGSF      648
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      592 KHYFVDSIAGDELMENSF---SPLA-----ATGKPLSRGCGKCHRPKMYTIOAKPSRL      641

QY      649 YLSCMGFPECRSAWMLPDS-----VL--EASRDS--VCPVQCPHFV      686
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      642 HCS-----HCDETYLTPQNGTKIKVELRCPDLDELYVMSSGSGKSKSYPLPCYNHPR      696

QY      687 YRLKLKFKRGLSLPTMPLEFVCC-----TGCDDTLREILDLRFSGGP      729
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      697 FR-DMKKKGWGCNECTHP---SCQHSLSMLIGQCEVCESGVYLVLDPTSGP      742

RESULT 9
TP3B_DROME STANDARD: PRT: 875 AA.
AC 096651; Q9W416;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase III beta (EC 5.99.1.2).
GN TOP3-BETA OR TOP3 OR CG3458.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RX NCBI_TaxID=7227;
RN [1]

```


RP SEQUENCE FROM N.A., FUNCTION, AND DEVELOPMENTAL STAGE.
RC TISSUE-Embryo;
RX MEDLINE=20102653; PubMed=10636841;
RA Wilson T.M., Chen A.D., Hsieh T.-S.;
RT "Cloning and characterization of Drosophila topoisomerase IIBeta.
RT Relaxation of hypernegatively supercoiled DNA.";
RL J. Biol. Chem. 275:1533-1540(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Adamatsis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencs P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fostler C., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sviderskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: WEAKLY RELAXES NEGATIVE SUPERCOILS AND DISPLAYS A
CC DISTINCT PREFERENCE FOR BINDING SINGLE-STRANDED DNA.
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE FIRST 6 HOURS OF
CC EMBRYONIC DEVELOPMENT, LEVELS DECLINE DURING LARVAL AND PUPAL
CC STAGES TO INCREASE AGAIN DURING ADULTHOOD.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AF090909; AAD13219.1; -
CC DR EMBL, AE003437; AAF46144.1; -
CC DR

DR FlyBase: FBgn0026015; Top3-beta.
DR InterPro: IPR003601; DNATopI_ATP_bind.
DR InterPro: IPR003602; DNATopI_DNA_bind.
DR InterPro: IPR000380; Pro_Topoisomerase.
DR InterPro: IPR002936; Toprim.
DR Pfam: PF01131; Topoisom_Bac; 1.
DR Pfam: PF01751; Toprim; 1.
DR PRINTS: PR00417; PRPISMASSET.
DR SMART: SM00437; TOP1AC; 1.
DR SMART: SM00436; TOP1BC; 1.
DR SMART: SM00493; TOP1IM; 1.
DR PROSITE: PS00396; TOPOISOMERASE_I_PROK; 1.
KW Isomerase; Topoisomerase; DNA-binding.
FT ACT_SITE 332 332 DNA_CLEAVAGE (BY SIMILARITY).
FT DOMAIN 819 859 GLT-RICH.
FT CONFLICT 747 747 V -> M (IN REF. 1).
SQ SEQUENCE 875 AA; 96973 MW; 3A26520C10AB6057 CRC64;
Query Match 19.6%; Score 1022; DB 1; Length 875;
Best Local Similarity 31.1%; Pred. No. 1.7e-58;
Matches 294; Conservative 123; Mismatches 376; Indels 152; Gaps 22;
QY 6 VRKYLVAERNDAAKGIADLLSNGRMRRRGSKFKRYEPDYHLYGO-VYTWVTSVSG 64
- : : | | | | : : | | | | : : | : : : | | | |
Db 1 MKSYLWAEKPSLAASLAGILSNCRGCTAKRGTCGCTHETGTFNMGSHVEMTSYVG 60
QY 65 HLAHDFOMQFRKWKQSNPLVLEAEIEKYCPENFVDIKKTLERETRCQALVITDCDR 124
:
Db 61 HVMSLDNRKKYKNCMDKDPDIOLEFCATEKKEKTNKQMMKFLAEARGCDVLYMLDCDK 120
QY 125 EGENIGFEII-----HVCAKAPNQLVLRARSETTPHAVRACBNLLEPDORVDAVDR 180
| | | | | | | | | | | | | | | | | | | | | | | |
Db 121 EGENICEVMDAYKAVNNVYSDVYTRAFHSALTEDIKKAMETLGHNPENAKSDVAR 180
QY 181 QEDLDRLGAATFRQTLRLORIPPEVLAEDLISGSCOPTLGFVVERFAKIAFVEIF 240
| | | | | | | | | | | | | | | | | | | | | | | |
Db 181 QEDLDRLGCAFTFRQTFKFDQRYGD-LDSSLISYPCQPTLGGCVKRRHDIQFKRESF 239
QY 241 HRIKVTDHDKDGLVEFMKKHRLFNHTRACLVLYOLCEDPMATVVEVRSKPKSKRRQAL 300
:
Db 240 WHIQLLAGQRP-VTELMAGRVFKKIDALMLNRVKEKKAYVESVASKAVYSKRDAL 297
QY 301 DYVELKILASRKRLINKETPRIRIAEKLTYGYSIPRTETNIFPRDLNVLVYQGTDPD 360
| | | | | | | | | | | | | | | | | | | | | | | |
Db 298 NYVELMNICSSGLGIGSPQAMQIBRLTYGYSIPRTETNQTNDLRAVLNLPSPA 357
QY 361 RWGAFQASILERG-PPIRNGKNSDOANPIHPKYTN-NLQGEORLYEFIVRHLAC 417
: : | | | | | | | | | | | | | | | | | | | | | |
Db 358 DFGEARSIL-GDIQTPRKQ-KDAGDHPITPKLGNRSDPDRDTRVVEFICRHFGT 414
QY 418 CSQDAQOQETTVEDIDQERVANGMLIARNLYLDVYPIYDMS-----DKLIPVEQSHF 473
:
Db 415 VSRDLKRYVTAKLSVMEFTFSCYASVLIDAGFTKVT--WSAFGRDEDPRPVQGTQV 471
QY 474 QPSVVEWDEGTSPKLTLEADLALMEKIGIGDATHAEIETIKRMVGLPRFL 533
:
Db 472 AINDVRLIESQTPPDVLTSELTITLMEENGIGDASIPVHINNICORNYNIENGKRLM 531
QY 534 PGHLGMLVEGYDSMGYEMSKPDLRAELADLKLICGKKDFVLRQOVQKYQVTEA 593
| | | | | | | | | | | | | | | | | | | | | | | |
Db 532 PTLGIYLVHGYQKIDDELPLPTMRTEVERMLTLIAGSANFQOVLRHAIKIFLKFME 591
QY 594 YAKKKKIDELALQYFGNGTELAQOEDITYPAMPETIRKCPQCNKMVLTKKNGGFYLSGM 653
: : | :
Db 592 V---KNIDSDALCEVVSFSLAESGAH-----SRGCKRMYKTYIQTRPARLDS- 639
QY 654 GFPECRSAVWLPDSVLASRDS-----VCPVCPRPYRRLK 692
| | : |
Db 640 ---HCDEFYALPIGNVAVYREFKCPRLDDFDLLARSTGVKSGSYFCYCYVHP----- 689
QY 693 FKRGSLPTMPLEFVCCIGCDQDLTRELTLRFSGGPRPASQPSGRLOANOSLNRMNSQ 752
:
Db 690 -----PFSDMPLHGGC-----NTCTNAN 707


```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Unwinding enzyme) (Swivelase).
GN TOPA OR TA0063.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;
OC Thermoplasma
NCBI_TaxID=2303;
OX NCBI_TaxID=2303;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Luppas A.N., Baumeister W.;
RT "The genome sequence of the thermophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL445063; CAC11211.1; -
DR InterPro: IPR003601; DNATopI_ATP_bind.
DR InterPro: IPR003602; DNATopI_DNA_bind.
DR InterPro: IPR000360; Pro_topoisomerase.
DR InterPro: IPR002936; Toprim.
DR Pfam: PF01131; Topoisom_bac; 1.
DR Pfam: PF01751; Toprim_1.
DR Pfam: PF01396; zi-C4_topoisom; 1.
DR PRINTS: PR00417; PRTPISMRASE1.
DR SMART: SM00437; TOP1AC; 1.
DR SMART: SM00436; TOP1BC; 1.
DR SMART: SM00493; TOPRIM; 1.
DR PROSITE: PS00396; TOPOISOMERASE_I_PROK; FALSE_NEG.
KW isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
KW Repeat; Complete proteome.
FT ZN_FING 611 638 C4-TYPE 1.
FT ZN_FING 673 700 C4-TYPE 2.
FT ZN_FING 719 744 C4-TYPE 3.
FT ACT_SITE 312 312 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 770 AA; 87667 MW; 75DA8DD7BC3B8A22 CRC64;
Query Match 12.5%; Score 653; DB 1; Length 770;
Best Local Similarity 26.5%; Pred. No. 1e-34;
Matches 205; Conservative 130; Mismatches 329; Indels 110; Gaps 22;

```

```

Db 113 EGGLEVEALDIIKEGEE--EIRAKFSALTKNKNEILDSFKNLIGVNSLADADARESID 170
Qy 185 LRIGAFTFRPOTLRLORIPEVLAEOILSYGSCQPTPLGVEFEKAIQAFPEIEHRK 244
Db 171 LINGSVLTFRFSVTGTR-----LGSFSLASGRVQPTPLAIVNDREREIQSFPERERYTIS 225
Qy 245 VTHDHKGIYEFPMKRRRLRFLNHNACVLVYOLQVEDPAAVVEVRSKSKMRQALDVE 304
Db 226 ITFD-KDGFKARIPENIKQDDTA-EKIVE-ALKGKGRSVSTYSKDDIRRPPESTTE 282
Qy 305 LEKLASRLKRLINAKETMRIAEKLYTGQYISYPTETNIFPPDNLTVLVEQGTDPDRMGA 364
Db 283 FLREASR-ICIMPTKAMSIENLYMRGLISYPRDNTVYRSINLKV-LKLENTAYSK 340
Qy 365 PAOSILERSGPTPRNGKSDQAHPIHP-KYNNLOGDEQRILEYVIRHFLACCSDDAQ 423
Db 341 YKIEIEFDRLISRSRIETTDHPPIYVPVDSPKQLGDRVYDLILRHFLSTLYRDGK 400
Qy 424 GQETTVIEDIAQERFVHGLMILARNLDVYVPYHMSDKILPYEQGSHQPSVEAVDG 483
Db 401 KYVAEALIVNGYTFKAGQHTTDRGWTETIYDIP-KDVIYLPETEGEDLKAIDMNTQRE 459
Qy 484 ETSPPKLLTEADLIALMEKHGIGTDATHAHEIETIKARMYGLTPDRRFLPGHGMCLVE 543
Db 460 ETRKPPRYDMSLSLKKNEELNLGTRKHIDIGLIERGFEIGNPVK---PTPLGMAFID 516
Qy 544 GYDSMGYEMSKPDILAEADLKLICDQKKDFYVLROQVYKQVFEIYAVAKAKLIDEA 603
Db 517 AVRSVNSHIADPEWTAKLEEDMDRIEKNEMSKNDYVEESKMLHEVLSHFLTAKYADKDI 576
Qy 604 LAQYFGNGTEL-----AQGEIYVAMPPIR-----K 630
Db 577 ITGGINAGIELGDCPFHEGKKNIVIRDRFTYVACED--PSCIKNFIRKNGSITLSDQK 634
Qy 631 CPQC-----NKD-----AYLTKRKNNGFYLSQ 652
Db 635 CPVCGELPMIKIRKQSPETKICIDPDSCSYNRENEDYECPADHGRVLVLRQSKRFLGCG 694
Qy 653 MGPEPCSAWLPDSVLEASRDSVCVCPHPYRRLKLFKRSLEPTPLER 706
Db 695 SNYPKCTVYPLP-QMGRITKTEGVCYGA-PLIALSRNGKMKFCFPMNOCEY 746
RESULT 12
TOP1_PYRAB STANDARD; PRT; 685 AA.
ID TOP1_PYRAB
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Unwinding enzyme) (Swivelase).
GN TOPA OR PAB1430.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
RT structure and evolution."
RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE

```


DR	SMART: SM00493; TOPRIM: 1.	DR	SMART: SM00396; TOPOISOMERASE_I_PROK; FALSE_NEG.	DR	SMART: SM00396; TOPOISOMERASE_I_PROK; FALSE_NEG.
DR	PROSITE, PS00396; TOPOISOMERASE; DNA-binding; Zinc-finger; Metal-binding;	DR	PROSITE, PS00396; TOPOISOMERASE; DNA-binding; Zinc-finger; Metal-binding;	DR	PROSITE, PS00396; TOPOISOMERASE; DNA-binding; Zinc-finger; Metal-binding;
KM	Isomerase, Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;	KM	Isomerase, Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;	KM	Isomerase, Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
KW	Repeat; Complete proteome.	KW	Repeat; Complete proteome.	KW	Repeat; Complete proteome.
FT	ZN_FING 606 634	FT	ZN_FING 606 634	FT	ZN_FING 606 634
FT	ZN_FING 653 678	FT	ZN_FING 653 678	FT	ZN_FING 653 678
FT	ACT_SITE 317 317	FT	ACT_SITE 317 317	FT	ACT_SITE 317 317
SO	SEQUENCE 686 AA: 78637 MW: 77590133CADDE944 CRC64;	SO	SEQUENCE 686 AA: 78637 MW: 77590133CADDE944 CRC64;	SO	SEQUENCE 686 AA: 78637 MW: 77590133CADDE944 CRC64;
Query Match	11.3%; Score 591; DB 1; Length 686;	Query Match	11.3%; Score 591; DB 1; Length 686;	Query Match	11.3%; Score 591; DB 1; Length 686;
Best Local Similarity	25.7%; Pred. No. 8.7e-31;	Best Local Similarity	25.7%; Pred. No. 8.7e-31;	Best Local Similarity	25.7%; Pred. No. 8.7e-31;
Matches 190; Conservative 138; Mismatches 307; Incls 104; Gaps 21		Matches 190; Conservative 138; Mismatches 307; Incls 104; Gaps 21		Matches 190; Conservative 138; Mismatches 307; Incls 104; Gaps 21	
QY	9 VLCAEKKDAKGIADLLSNGMRREGISLKNKIYEDYHLGYGVNTVMWTSVSGHLI- 67	QY	9 VLCAEKKDAKGIADLLSNGMRREGISLKNKIYEDYHLGYGVNTVMWTSVSGHLI- 67	QY	9 VLCAEKKDAKGIADLLSNGMRREGISLKNKIYEDYHLGYGVNTVMWTSVSGHLI- 67
DB	2 ILIIAEKPNVARKKIGALSERRPIK---SLFGVPY---YEIFREGKTLIVASAVGHLYG 55	DB	2 ILIIAEKPNVARKKIGALSERRPIK---SLFGVPY---YEIFREGKTLIVASAVGHLYG 55	DB	2 ILIIAEKPNVARKKIGALSERRPIK---SLFGVPY---YEIFREGKTLIVASAVGHLYG 55
QY	68 -----ADFQWQFRRKQSCNPLVFEAFIEKYCPENFEDIKTLERETROGQALVIMWD 121	QY	68 -----ADFQWQFRRKQSCNPLVFEAFIEKYCPENFEDIKTLERETROGQALVIMWD 121	QY	68 -----ADFQWQFRRKQSCNPLVFEAFIEKYCPENFEDIKTLERETROGQALVIMWD 121
DB	56 LAPKRDVGYPIFDIEW-----VPYIIAEKGEYAREYIKALSYLAKRYE---FIVACD 107	DB	56 LAPKRDVGYPIFDIEW-----VPYIIAEKGEYAREYIKALSYLAKRYE---FIVACD 107	DB	56 LAPKRDVGYPIFDIEW-----VPYIIAEKGEYAREYIKALSYLAKRYE---FIVACD 107
QY	122 CDREGENIGFELIIHCKAVKPMIQLVLRASFETPHAVTACENTLEP--DQVSDAVDV 179	QY	122 CDREGENIGFELIIHCKAVKPMIQLVLRASFETPHAVTACENTLEP--DQVSDAVDV 179	QY	122 CDREGENIGFELIIHCKAVKPMIQLVLRASFETPHAVTACENTLEP--DQVSDAVDV 179
DB	108 YDTEGEVIGYATLKYACGYDPRV-AKRMKFSALTRDLDLMARNL-EPIINGMAGIA 165	DB	108 YDTEGEVIGYATLKYACGYDPRV-AKRMKFSALTRDLDLMARNL-EPIINGMAGIA 165	DB	108 YDTEGEVIGYATLKYACGYDPRV-AKRMKFSALTRDLDLMARNL-EPIINGMAGIA 165
QY	180 ROELDLRIGAFTRPOTLRQRIPEVLAESQLISGSCFPPLGLGVEFEFKAIQAFVPEI 239	QY	180 ROELDLRIGAFTRPOTLRQRIPEVLAESQLISGSCFPPLGLGVEFEFKAIQAFVPEI 239	QY	180 ROELDLRIGAFTRPOTLRQRIPEVLAESQLISGSCFPPLGLGVEFEFKAIQAFVPEI 239
DB	166 RHILDMYGVNLSRALTLHAIKKASKWV--VLSTGRVGQPLTLKFLVEREREIOSFVRPR 222	DB	166 RHILDMYGVNLSRALTLHAIKKASKWV--VLSTGRVGQPLTLKFLVEREREIOSFVRPR 222	DB	166 RHILDMYGVNLSRALTLHAIKKASKWV--VLSTGRVGQPLTLKFLVEREREIOSFVRPR 222
QY	240 FHRIVTVDHDKGCIYEPMKKRHRHLNHTACLVLYQCVADPMATVVEVS----- 289	QY	240 FHRIVTVDHDKGCIYEPMKKRHRHLNHTACLVLYQCVADPMATVVEVS----- 289	QY	240 FHRIVTVDHDKGCIYEPMKKRHRHLNHTACLVLYQCVADPMATVVEVS----- 289
DB	223 YWVILIEKNGQKPTANEEKDIMEE-----EEGRIVLEVAKKSIIPRNSVNEI 271	DB	223 YWVILIEKNGQKPTANEEKDIMEE-----EEGRIVLEVAKKSIIPRNSVNEI 271	DB	223 YWVILIEKNGQKPTANEEKDIMEE-----EEGRIVLEVAKKSIIPRNSVNEI 271
QY	290 KPKSKWRPDALDTVELEKSLASRKLRIINAKETMRIAEKLTYOGIYSYPRFTNIFPDLNL 349	QY	290 KPKSKWRPDALDTVELEKSLASRKLRIINAKETMRIAEKLTYOGIYSYPRFTNIFPDLNL 349	QY	290 KPKSKWRPDALDTVELEKSLASRKLRIINAKETMRIAEKLTYOGIYSYPRFTNIFPDLNL 349
DB	272 KRQKTRPHRPELGLQRRAYSAFGSPSKKTIIDIASLIEKGFSSYPRFESQKLPNNINF 331	DB	272 KRQKTRPHRPELGLQRRAYSAFGSPSKKTIIDIASLIEKGFSSYPRFESQKLPNNINF 331	DB	272 KRQKTRPHRPELGLQRRAYSAFGSPSKKTIIDIASLIEKGFSSYPRFESQKLPNNINF 331
QY	350 TVLVEQTPRDPKMGAPAOISILERGGPTPRNGKSDOANRPINPT--KYTNLQGDQRL 406	QY	350 TVLVEQTPRDPKMGAPAOISILERGGPTPRNGKSDOANRPINPT--KYTNLQGDQRL 406	QY	350 TVLVEQTPRDPKMGAPAOISILERGGPTPRNGKSDOANRPINPT--KYTNLQGDQRL 406
DB	332 RMIIQNLAKMPQRYRYAHILLGLPELKPVEGKKEDPAHAIYPTGEIRPRGDLTKDEEKL 391	DB	332 RMIIQNLAKMPQRYRYAHILLGLPELKPVEGKKEDPAHAIYPTGEIRPRGDLTKDEEKL 391	DB	332 RMIIQNLAKMPQRYRYAHILLGLPELKPVEGKKEDPAHAIYPTGEIRPRGDLTKDEEKL 391
QY	407 YEFIRHFLACSSODAOGETVEIDIAQERRVANGMLTANRYLDVY-PYDHWMSKILRP 465	QY	407 YEFIRHFLACSSODAOGETVEIDIAQERRVANGMLTANRYLDVY-PYDHWMSKILRP 465	QY	407 YEFIRHFLACSSODAOGETVEIDIAQERRVANGMLTANRYLDVY-PYDHWMSKILRP 465
DB	392 YDMIRRLFAEMERAIRSVKVTIRAGPHKFFLSGGRTVKKGMWLSVYKRYKFFEEVTLRP 451	DB	392 YDMIRRLFAEMERAIRSVKVTIRAGPHKFFLSGGRTVKKGMWLSVYKRYKFFEEVTLRP 451	DB	392 YDMIRRLFAEMERAIRSVKVTIRAGPHKFFLSGGRTVKKGMWLSVYKRYKFFEEVTLRP 451
QY	466 VYEGQSHFQPSVFEWVDGTSPPKLLTEADLLALMEKHGIGDPAHEHIEVITKARMYVG 525	QY	466 VYEGQSHFQPSVFEWVDGTSPPKLLTEADLLALMEKHGIGDPAHEHIEVITKARMYVG 525	QY	466 VYEGQSHFQPSVFEWVDGTSPPKLLTEADLLALMEKHGIGDPAHEHIEVITKARMYVG 525
DB	452 EEFIEERIKVIOYKKEKKTKTPARKSPAAVYIKKMDLGLGKATRAQILLEIILYQGYIE 511	DB	452 EEFIEERIKVIOYKKEKKTKTPARKSPAAVYIKKMDLGLGKATRAQILLEIILYQGYIE 511	DB	452 EEFIEERIKVIOYKKEKKTKTPARKSPAAVYIKKMDLGLGKATRAQILLEIILYQGYIE 511
QY	536 LTPDKRFLPGHLMGLVEGYDSMGYEMSKPDLRAELEADLKLIDCKKDKFVYLROQYOK 585	QY	536 LTPDKRFLPGHLMGLVEGYDSMGYEMSKPDLRAELEADLKLIDCKKDKFVYLROQYOK 585	QY	536 LTPDKRFLPGHLMGLVEGYDSMGYEMSKPDLRAELEADLKLIDCKKDKFVYLROQYOK 585
DB	512 GKKSIIKVP--LGMKVIETLEKYVPEIISVELTRFEKKKMLIMEGRLTKEEVIEAK				

DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Unwisting enzyme) (Swivelase).
GN TOPA OR MO1652.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae;
OC Methanococcus.
CX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RX Bult C.J., White O., Olsen G.J., Zhou L., Fleischman R.D.,
RA Sultun G.G., Blake J.A., Fitzgeraid L.M., Clayton R.A., Gocayne J.D.,
RA Kerepasek A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Kerepasek R., Kirkness E.F., Weinstock K.G., Merriell J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Brodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC SIMILARITY).
CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67605; AAB9673.1; -;
DR HSSP; P06612; IECL.
DR TIGR; M11652; -;
DR InterPro; IPR0033601; DNATopI_ATP_bind.
DR InterPro; IPR003602; DNATopI_DNA_bind.
DR InterPro; IPR000380; Pro_topoisomerase.
DR InterPro; IPR002936; Toprim.
DR Pfam; PF01131; Toprim_bac; 1.
DR Pfam; PF01751; Toprim; 1.
DR Pfam; PF01396; Zf-Ca_topoisom; 3.
DR PRINTS; PR00417; PRTPISMRASEI.
DR SMART; SM00437; TOP1AC; 1.
DR SMART; SM00436; TOP1BC; 1.
DR SMART; SM00493; TOPRIM; 1.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW Isomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
KW Repeat; Complete proteome.
FT ZN_FING 600 626 C4-TYPE 1.
FT ZN_FING 680 706 C4-TYPE 2.
FT ZN_FING 721 747 C4-TYPE 3.
FT ACT_SITE 315 315 DNA CLEAVAGE (BY SIMILARITY).
FT SEQUENCE 761 AA: 87832 MW: 29935753E202D82 CRC64;

```

Query Match      11.0%: Score 573; DB 1: Length 761;
Best Local Similarity 27.1%: Pred. No. 1.5e-29;
Matches 196; Conservative 115; Mismatches 337; Indels 74; Gaps 19;

Oy      6 VRKVCVAEKNDAAKGIDLLSNGRRRREGSKFNKIYEFQYHLYGQNTVMVTMSVSGH 65
      : : : : | | | | : : : : : : : : : : : : : : : : : :
Db      4 IMTALITKEKSPVAKKINNALGAKAKKSIDGVPY-----YELERDQKIIIVASAVGH 55

```

```

QY 66 LLAHDFQMFRRKWCSCNPLVLEAEIEKYCPBNEVDIK-----KTLER 108
D 56 L-----FTLVEKENKEFGFYVPYDIKWVASVDKGEYVKKYIKALK 98
QY 109 ETRQCALVITWDDREGENIGFEIHHYCKAVKPNLOYLRARFSEITTHAVTAENLLE 168
D 99 LSKDADEFYIATDWIDEBELIGYHALKYCCGE--KAKRMFSSLTKEIVRAENDE 155
QY 169 PDORVSDAVDYROELDLIGAFTFRQTLRLQRIPEVLAEOILSYGSCOPTLGFVVER 228
D 156 IDYGLVDAGESRHILDMVEGINLSR----ALMNAIRAVNRKMTSGVGVGAFALFLER 211
QY 229 FKALQAFPELPFHRKTYHDKDGVFEENKRNHLEFNTACILVYOLCEVEDMATVVEVR 288
D 212 ELEIKKFFPKPYWYEAU--LKDNLKATN-EKEKFWNEKEAKNYEIKDKDESAKVEIK 268
QY 289 SKPSKMPR-QALDTVELEKTLASRKLRIKAKETMRIAEKLTQGYISYPTETNFPDIL 347
D 269 -KTKRKRLKPLPFDLGTIQRREAYSFYKISPKETOIAOKLENLISPTSSOKLPDR 327
QY 348 N-LTVLVEOQTTPDRMGAFASQILERGGTPRNGKNSDOAHPIHPKY-TNNLOGDEOR 405
D 328 KYLEDILNIITKHPYVGKWAERILKE-NLKPEVEKKEDPAHAIHVIDIPEELSEKEKE 386
QY 406 LYEPVIRHFLACCSODAGQETTYEIDIAQERFVAHGMILARNLDVYPRDHMSDKLIP 465
D 387 IYDLIARTTLAFAWDMNAREYLNKIDIKGEKFKLSGSRTRYEGHELYEPKDEILP 446
QY 466 VYEOGSHRQPTSEVMEVDSTSPKLLTEADLIAMEKHGIGDATHAHIEIKARMYVG 525
D 447 PLKKNIDIKVKKITITRKETOPPKRYTVAASITKELEKGLDTKATRAIIDLKIKGYV- 505
QY 526 LTPKRELPLGHLNGGLVGYDSMGEMSKPDLRAELADLILICDGGKKEVULRQOYOK 585
D 506 -IDGSLKLVTDLGISVIEFTLRKCFEIIIDEKMTDLLEKLEIQRKIKKDDVDEAEKR 564
QY 586 YKQFIEVNAKAKLDELALAYFGNGTSLAQOEDTTPMPERIKPCPCQNDMYLTKTKN 645
D 565 LKRLIE---FKKREEDIGIY-----LIKLNATNKAIVGCPGCGDILLIRKK 614
QY 646 GGFYLSGWFPECRSAVWLPSVLEASRDSVCPQHPYVRLKLRKRSLSLPTMLE 705
D 615 GRF-VGCSNVECDYKYSPLPKG-RIKIPNKVDACK-SPI--LKIGREICINCEPLK 669
QY 706 FV 707
D 670 QV 671

```

```

RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL delta: functional analysis and comparative genomics.";
CC J. Bacteriol. 179:7135-7155(1997).
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC SIMILARITY).
CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000921; AAB86097.1; -.
DR HSSP; P06612; 1ECL.
DR InterPro; IPR003601; DNATOP1_ATP_bind.
DR InterPro; IPR003602; DNATOP1_DNA_bind.
DR InterPro; IPR000380; Pro.topoisomerase.
DR InterPro; IPR002936; TopIim.
DR Pfam; PF01131; TopIisom_bac; 1.
DR Pfam; PF01751; TopIim; 1.
DR Pfam; PF01396; zf-C4_TopoIisom; 2.
DR PRINTS; PR00417; PRTPISMRASEI.
DR SMART; SM00437; TOP1AC; 1.
DR SMART; SM00436; TOP1BC; 1.
DR SMART; SM00493; TOP1IM; 1.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; FALSE_NEG.
DR Isomerase; Topoisomerase; DNA-binding; zinc-finger; Metal-binding;
KW Repeat; Complete proteome.
FT ZN_FING 598 626 C4-TYPE 1.
FT ZN_FING 680 706 C4-TYPE 2.
FT ACP_SITE 320 320 DNA_CLEAVAGE (BY SIMILARITY).
FT SEQUENCE 718 AA; 88507CAED59C0307 CRC64;
SQ
Query Match 10.8%; Score 563; DB 1; Length 718;
Best Local Similarity 25.9%; Pred. No. 6e-29;
Matches 185; Conservative 120; Mismatches 306; Indels 102; Gaps 20;

```

```
QY 338 TETNIPRLNLTVLVEOOTPDPRMGAFASQILRGGPTPRNGNKSQDAHPPIHPT-KYT 396
| : | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 TSSQKLPESIGYEKILNLAKNRFGVHIERL--RGPLKHGKKEDDAHPATHPTGLP 380
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 397 NNIGDEORLYEFIVRHFLACCSQDAOGQETVEIDIAQERYVAHGIMILARNYLDVYPY 456
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 SELSKDEKRYDLIVHRFISVFGEDAILQTMKVELEIGEEFSEFSRKRYSKAGWMESYPY 440
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 457 DHMSDKILPYEGSHFQPSVTEMVDGETSPPKLLTEADLIALMEKHGIGTDATHAEHIE 516
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 441 TKMEDEEFPISGDSLAVRSVADEREKPPARYNBASLIRELERGLGKSTRADIIA 500
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 517 TIKARYVGLTPDKRFLPGHLMGLVEGYDSMGYEMSKPDLRAELADLKLICGKKDKF 576
: | : | : : | : : : : : : : : : : : : : : : : : : : :
Db 501 KLYDRKYI---EGKTIKRVSPLENIIDTLTRYCEKIISEELTRQFERELDIMRGK---- 553
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 577 VVLRQOVQKIKQVFIENAVAKAKKLDALAOYFGNGTELAQO-----EDITYPAMPEP--IR 629
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 554 -----ISKDRVIDEAITFEVRSILSDIEENLRDIGKELYRAYODSRVYG 596
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 630 KCPQCNKDWLTK--TKKNGGFYLSMGFPPECRSVWLPDSVLEASRDSVCPYC 681
: | : | : : | : : : : : : : : : : : : : : : : : : : :
Db 597 ECPACGGKLVTKYSPNRNSTFVGCCSTPDCRTVYSLPRG---ASVLKSLCEKC 646
: | : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: October 11, 2002, 15:26:14
Job time : 18 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 13, 2002, 13:48:45 : Search time 6334.08 Seconds
(without alignments)
9822.203 Million cell updates/sec

Title: US-09-744-125-1

Perfect score: 2973

Sequence: 1 ggaaccgagcgcgcagatggc.....tagagaagcttggagtcgac 2973

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
------------	-------------	-------	--------	----	----	-------------

1	2949.6	99.2	3755	9	HSU43431	U43431 Human DNA t
2	2233.2	75.1	3741	10	AB006074	AB006074 Mus muscu
C	594.6	23.1	132486	9	HSF10178	AL035367 Homo sapi
C	446	20.0	3891	3	AF255733	AF255733 Drosophill
5	434.8	15.0	88718	2	AC097545	AC097545 Rattus no
C	434.8	14.6	203117	10	AL596215	AL596215 Mouse DNA
C	433.4	14.6	213913	2	AC069534	AC069534 Mus muscu
9	433.4	14.6	87835	3	AC005428	AC005428 Drosophill
10	433.4	14.6	157766	3	AC099017	AC099017 Drosophill
11	429.4	14.4	299449	3	AE003663	AE003663 Drosophill
12	422	14.2	73907	2	AC018257	AC018257 Drosophill
13	270.6	9.1	2422	3	AF057032	AF057032 Caenorhab
14	270.6	9.1	2784	9	AF017146	AF017146 Homo sapi
15	270.6	9.1	2793	9	AF053082	AF053082 Homo sapi
16	269	9.0	3133	9	BC002432	BC002432 Homo sapi
17	262	8.8	2680	9	AF125216	AF125216 Homo sapi
18	257.8	8.7	218329	8	CNS07ECB	AL590444 Chromosom
19	253.2	8.5	1869	8	AF126287	AF126287 Schistosac
20	240.6	8.1	110000	2	LMFICHR36_26	Continuation (27 o
21	238.4	8.0	2979	3	AF099909	AF099909 Drosophill
22	233.6	7.9	3267	3	AF061625	AF061625 Drosophill
23	195.2	6.6	2780	10	AB013603	AB013603 Mus muscu
24	194.8	6.6	3543	10	AB045324	AB045324 Mus muscu
25	177.4	6.0	2390	6	AX073290	AX073290 Sequence
26	172	5.8	37619	2	AC091210	AC091210 Drosophill
C	147	4.9	245844	2	AC021094	AC021094 Homo sapi
27	136.2	4.6	169364	3	CEY36A3A	AL132860 Caenorhab
28	136.2	4.6	2435	8	YSCTOP3	M24339 Yeast (S.ce
29	136.2	4.6	2471	6	AX073076	AX073076 Sequence
30	136.2	4.6	41454	8	YSC18083	U19027 Saccharomyc
31	133	4.5	71516	2	AC014220	AC014220 Drosophill
C	133	4.5	305018	3	AE003437	AE003437 Drosophill
32	115	3.9	207184	2	AC020664	AC020664 Homo sapi
C	105.8	3.6	37619	2	AC091210	AC091210 Drosophill
C	95	3.2	10373	2	AC023710	AC023710 Drosophill
36	94.8	3.2	13114	1	AE009905	AE009905 Pyrobacul
C	89.8	3.0	37852	8	SPBC1665	AL023554 S.pombe c
37	77.6	2.6	251700	1	AP000062	AP000062 Aeropyrum
38	77	2.6	11015	6	AX280003	AX280003 Sequence
39	77	2.6	11015	6	AX281190	AX281190 Sequence
40	77	2.6	11015	6	AX356459	AX356459 Sequence
41	76.8	2.6	344000	1	TAC1D1	AL445063 Thermopla
C	42	2.5	165027	3	CEY48C3A	AL117203 Caenorhab
43	74.8	2.5	88718	2	AC097545	AC097545 Rattus no
44	74	2.4	958	33	AC085945	AC085945 Giardia 1
45	70.2	2.4				

ALIGNMENTS

RESULT 1
LOCUS HSU43431 3755 bp mRNA linear PRI 12-JUL-1996
DEFINITION Human DNA topoisomerase III mRNA, complete cds.
ACCESSION U43431
VERSION U43431.1 GI:1292911
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 3755)
AUTHORS Hanai,R., Caron,P.R. and Wang,J.C.
TITLE Human TOP3: a single-copy gene encoding DNA topoisomerase III
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (8), 3653-3657 (1996)
MEDLINE 96195027
REFERENCE 2 (bases 1 to 3755)
AUTHORS Hanai,R. and Wang,J.C.
TITLE Direct Submision
JOURNAL Submitted (15-DEC-1995) Ryo Hanai, Molecular and Cellular Biology,
Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA
FEATURES
source
1..3755

/organism="Homo sapiens"
 /db.xref="taxon:9606"
 /chromosome="17"
 /map=17p11.2-12"
 /cell_line="Jurkat T-cell line, ATCC E6-1"
 178. 3741
 /gene="TOP3"
 178. 3183
 /gene="TOP3"
 /note="putative; uses an alternative initiation codon"
 /codon_start=-1
 /product="DNA topoisomerase III"
 /protein_id="AA803694.1"
 /db.xref="GI:1292912"
 /translation="MIFPVARVLYLNRARRPEDPAFASRAMENALGVRKVLCAVKEAD
 AAKIADLNSGMRNRREGISFKNIYEDYLLYCONYMTALSVGLHLANDROMO
 RKMSQVLLFPEELIKESQKPEHFVTKTLETRRPOCALVITWDCEBENIGPELI
 HKSCAKVPLVQGLVLRARFESPTPHARVTCENLTPRQVSPANDVARELDLIGAR
 TRPTVLLAQILPPEVLYAEOLISTGSCOPTLGFVYBRKADIDAEPELFIHKIKYTHDH
 KDGLVNRKRLHFNHTKVLVYGLCVSDPDAIVFNRKSKSWRRQALDYLELR
 LASKRLINAKETKRLHFNHTKVLVYGLCVSDPDAIVFNRKSKSWRRQALDYLELR
 AGLTLEGGGQTPPNKNGSDQAHNPHTVYTNNGDGRVLEFIVRFLACGSDQAQ
 GOETTVLEIDIAQRFGNAKGMILANRYLDLVYEDMSDKILPVEQSGSHQPSVEYEA
 GIVSGPRLKLTLDLALMEKHGIGDATHAHIEFIAKRWVGLTPKRFVPLFLGKMG
 GIVGYSQSMYEMSKDPLRAELEADLKICDGKKRKFVLRQVQKQKQVLEPAVAKA
 KILDEALAYQVSLGTELAQDDEIVYAMPPIKSCQCKKMDVATIKTKNGYSLSCMEF
 PECSAAVLPDVSYLEASRDSQVPCVQPHPIVRLKLRKRGSILPTTMLEVCCTGGC
 DDITLRELIDRFGGPPRRASQSPGRLOANQSLNRNDNSQNDOPADSRQSSKALAO
 LPPATGAAGENSATVQCCQDEAVLLIVRKEGPRGRQFKRCNGSCNFFLAMDSPNPA
 GGPVALPARGLSAQCCPPQGIHLGGFNGPDSGSGSCISQSPSVTRVQKGDGPN
 KGRQFHPCAKPRDQCGFQWVNDHAGPNTDSCAPSWMTDRCRTLESEARSKPRASSS
 MGSTAKPKKSCILQDQGHTRFCQNS"

BASE COUNT	896 a	1020 c	1061 g	778 t
ORIGIN	<p> 253. .3183 /gene="Top3" /note="type I DNA topoisomerase, belonging to E. coli topoisomerase I/yeast topoisomerase III subfamily" /codon_start=1 /product="DNA topoisomerase III" /protein_id="AAB03695.1" /db_xref="GI:1292913" /translation="MEMALRGKRVLYLVAEKNDAAKSIADLLSNGRRRREGLSKFK IYEDYDHLVGGWVTWMTSVSGHLADFQWQFKWQSCNLYLFEAEIEYCEPENV DIKTLTERETKOCALYIWTDCRGEENIGEIIHWKAKVPIHQVLRARSETTPHA VIRTCENLTEDDQVSADVVRQELDRIGAEVFRFOTLRQRFEPVLAEOLISYSS COPTFLVEVERFAKIOAFVEIPIHRTKIVHDHGDVIEFNMKHRLFNFTACTLYG LCVDFPMATVEVERSKRSKRKRQALDITVELEKSIKRIINAKEMTIRAEKLYGVY ISYRTEINIPRDLNLTLYVEOOTPPRNKCAPOSILBERGPTPRNGKNSDAAHPT HPVITPNNLQGEORLVEYIVRHLACCSODAQOETTVLEIDIAQEFVAVHGLTAR NLYDVPYDHWMSKILPVYEQGSHFOSTYEMDGETSPVLTLEADILAMEKHGIG LTDAKHAEIEIKARMYVGLTPDKRFLPGHIGMLVEGYSMDGEMSKPDLRAELAD KIDLCGKDEKFEVYLROOVOKVOVFLTEAVAKAKLEALAOYNGNTELAOQDILY AMPRIKPCQCNKMDILTKKNGGFTLSCGPECRSAVLPVSGTVAESRDSQCV CQHPVYLRKLKTKRGSLLPTMPLEPTCCIGGCDDTLREILLDRFSGCPPRASPSCR LQAQOSTLRMDSQHPQPADSRQTSKALAQTLPTPTAAGESNYSVCNGQEVALLT VFKSGPNRGRQFSGKNGSCNFMFLMADSPNGAGPPALAYRLGASLGGPFGHIL GFEFNPGRDGSFGSCSCQSPSVYTRQKDGKNGRQFHTCAKREDCGCFQWVDN TAPTSGAPSWTGDGRGRTLSEARSKNRPASSSDMGSTAKKPKRCSLCHQGHTRPC PQR"</p>			
variation	<p> 1964 /gene="Top3" /replace="a" 3354 /gene="Top3" /replace="c" 3736. .3741 /gene="Top3" </p>			
variation	<p> polyA_signal </p>			

OY	5	CCGAGCTCGACATGCGCCCTCCGAGAGCGGTGGGAAAAGCTCTGTGTGGCCCGAAAAAAG	64
Db	248	CCGCCATGTGAATGATGGCCCTCCGAGCGCTGGGAAAAGTCTGTGTGGCCCGAAAAAAG	307
OY	65	ACGGGCGCAAGGGGATGCGCGACCTGCTGTCAAAACGGTGCATGAGCGCGAGAGAAGAC	124
Db	308	ACGGCGCCAAAGGGATGCGCGACCTGCTGTCAAAACGGTGCATGAGCGCGAGAGAAGAC	367
OY	125	TTTCAAAATTCACAGAATCTATGAATTTGATTTATCATCTGTATGGCCAGAAATGTTACCA	184
Db	368	TTTCAAAATTCACAGAATCTATGAATTTGATTTATCATCTGTATGGCCAGAAATGTTACCA	427
OY	185	TGTGAATGACTTCAGTTTCTGAGCAATTTACTGGCTCATGATTTCCAGATGACGATTTGCA	244
Db	428	TGTGAATGACTTCAGTTTCTGAGCAATTTACTGGCTCATGATTTCCAGATGACGATTTGCA	487
OY	245	AATGGCAGAGTGGCAACCCCTTCTCTCTTTGAACGAGAAATTTGAAAGTACTGCCAG	304
Db	488	AATGGCAGAGTGGCAACCCCTTCTCTCTTTGAACGAGAAATTTGAAAGTACTGCCAG	547
OY	305	AGAAATTTTGTAGACATCAAGAAAACTTTGGAGAGAGACTGCGCAGTCCAGGCTCTGG	364
Db	548	AGAAATTTTGTAGACATCAAGAAAACTTTGGAGAGAGACTGCGCAGTCCAGGCTCTGG	607
OY	365	TGATCTGGACACTGACTGTGTATGAGAGGCGCAAAACATCGGGTTTGAATTTATCCAGCTGT	424
Db	608	TGATCTGGACACTGACTGTGTATGAGAGGCGCAAAACATCGGGTTTGAATTTATCCAGCTGT	667
OY	425	GTAAGCGCTAAACCCCAATCTGAGGTGTGGGACCCGATTCCTGTGATCCACACCC	484
Db	668	GTAAGCGCTTAABCCCAATCTGAGGTGTGGGACCCGATTCCTGTGATCCACACCC	727
OY	485	ATGCGCTAGAGCAGCTTGTGAAAACCCTGCACGAGCCTATCAAGAGGATGAGCGATGCTG	544
Db	728	ATGCGCTAGAGCAGCTTGTGAAAACCCTGCACGAGCCTATCAAGAGGATGAGCGATGCTG	787
OY	545	TGATGTGAGGCAGAGACTGGACTTGAGGATTTGGAGCTGCTTTACTAGTTCAGAGACC	604
Db	788	TGATGTGAGGCAGAGACTGGACTTGAGGATTTGGAGCTGCTTTACTAGTTCAGAGACC	847
OY	605	TGCGGCTTCAGAGGATTTTTCCTGAGGTGCTGGCAGACGCTCATCAGTTCAGCGACGT	664
Db	848	TGCGGCTTCAGAGGATTTTTCCTGAGGTGCTGGCAGACGCTCATCAGTTCAGCGACGT	907
OY	665	GCCAGTTTCCCAACACTGGGCTTGTGTGTGAGGCGTTCAACGCCATTCAGGCTTTTATC	724
Db	908	GCCAGTTTCCCAACACTGGGCTTGTGTGTGAGGCGTTCAACGCCATTCAGGCTTTTATC	967
OY	725	CAGAAATCTTCACAGATTTAAAGTAATCTATGACCAACAAGATGATCTGTGAATTTCA	784
Db	968	CAGAAATCTTCACAGATTTAAAGTAATCTATGACCAACAAGATGATCTGTGAATTTCA	1027
OY	785	ACTGGAAAAGGCATCGACTTTTAACCAACGCGCTGCCATGTCATTCAGTTGTGTG	844
Db	1028	ACTGGAAAAGGCATCGACTTTTAACCAACGCGCTGCCATGTCATTCAGTTGTGTG	1087
OY	845	TGGAGATATCCATGGCAACTGTGGTAGAGTGCATCTAAAGCCAAAGACAGTGGCGGC	904
Db	1088	TGGAGATATCCATGGCAACTGTGGTAGAGTGCATCTAAAGCCAAAGACAGTGGCGGC	1147
OY	905	CTCAAGCCTTGGACACTGTGGAGCTTGAGAGCTTGCGCTTCTCGAAAAGTTGAGAAATTAATG	964
Db	1148	CTCAAGCCTTGGACACTGTGGAGCTTGAGAGCTTGCGCTTCTCGAAAAGTTGAGAAATTAATG	1207
OY	965	CTAAAGAAACCATGAGGATTTGCTGAGAAAGTCTTACACTCAAGGGTACATCACTATCTCCC	1024
Db	1208	CTAAAGAAACCATGAGGATTTGCTGAGAAAGTCTTACACTCAAGGGTACATCACTATCTCCC	1267
OY	1025	GAACAGAAACCAACATTTTTCGCCAGAGACTTAAACTGACGGGTGTGGTGAACAGCAGA	1084
Db	1268	GAACAGAAACCAACATTTTTCGCCAGAGACTTAAACTGACGGGTGTGGTGAACAGCAGA	1327
OY	1085	CCCCCGATCCACGCTGGGGGGGCTTTTGCCAGAGCATTCATAGAGCGGGGTGTGCCACCC	1144

```

|||||
Db 1328 CCCCGATCAGCGCTGGGGGGCTTTGGCCAGAGCATCTAGAGCGGGGGTCCACCC 1387
Oy 1145 CACGCAATGGGAACAAGTCTGACCAAGCTCACCCTCCCATTCACCCACCACCAATACACA 1204
Db 1388 CACGCAATGGGAACAAGTCTGACCAAGCTCACCCTCCCATTCACCCACCACCAATACACA 1447
Oy 1205 ACAACTTACAGGAGATGAAACAGCCACTGACAGTTTATTTGTCGCAATTTCTGGCTT 1264
Db 1448 ACACTTACAGGAGATGAAACAGCCACTGACAGTTTATTTGTCGCAATTTCTGGCTT 1507
Oy 1265 GCTGCTCCAGAGATGCTCAAGGGGAGAGAGACCAAGTGGAGATGACATCGCTCAGAGAC 1324
Db 1508 GCTGCTCCAGAGATGCTCAAGGGGAGAGAGACCAAGTGGAGATGACATCGCTCAGAGAC 1567
Oy 1325 GCTTTTGGCCCATGGCTCTCATGATTTCTGGCCGAAACTATCTGATGTGTATCCATPAG 1384
Db 1568 GCTTTTGGCCCATGGCTCTCATGATTTCTGGCCGAAACTATCTGATGTGTATCCATPAG 1627
Oy 1385 ATCACTGAGATGACAAGATCCTCCTGTATGAGCAAGGATCCACTTTCAGCCAGCA 1444
Db 1628 ATCACTGAGATGACAAGATCCTCCTGTATGAGCAAGGATCCACTTTCAGCCAGCA 1687
Oy 1445 CCGTGGAGATGTGAGAGGGGAGAGACCCACCAAGCTGCTCACCGAGCCGACTCA 1504
Db 1688 CCGTGGAGATGTGAGAGGGGAGAGACCCACCAAGCTGCTCACCGAGCCGACTCA 1747
Oy 1505 TTGGCCCTCATGAGAGACATGGCATTTGTATGATTCATGGGCTATGAATGTCTAAACCTGACC 1564
Db 1748 TTGGCCCTCATGAGAGACATGGCATTTGTATGATTCATGGGCTATGAATGTCTAAACCTGACC 1807
Oy 1565 CCATCAAGACCCGAGATGTGAGGGGCTCACCCAGACAAAGGGTTCCTCCCTGGGGCACC 1624
Db 1808 CCATCAAGACCCGAGATGTGAGGGGCTCACCCAGACAAAGGGTTCCTCCCTGGGGCACC 1867
Oy 1625 TGGGCTGAGACTTGTGGAAGGTTATGATTCATGGGCTATGAATGTCTAAACCTGACC 1684
Db 1868 TGGGCTGAGACTTGTGGAAGGTTATGATTCATGGGCTATGAATGTCTAAACCTGACC 1927
Oy 1685 TCGGGGCTCAACTGGAAGCTGATCTGATGATCTGATGAGCAAAAGGACAAATTTTG 1744
Db 1928 TCGGGGCTCAACTGGAAGCTGATCTGATGATCTGATGAGCAAAAGGACAAATTTTG 1987
Oy 1745 TGGTTTAAAGGAGCAAGTGCAGAAATACAGAGGTTTTCATTGAAGCGGTGGCTAAAG 1804
Db 1988 TGGTTTAAAGGAGCAAGTGCAGAAATACAGAGGTTTTCATTGAAGCGGTGGCTAAAG 2047
Oy 1805 CAAAGAAATTGAGAGAGGCTTGGCCAGTACTTTGGGAATGGGACAGAGTTGGCCAGC 1864
Db 2048 CAAAGAAATTGAGAGAGGCTTGGCCAGTACTTTGGGAATGGGACAGAGTTGGCCAGC 2107
Oy 1865 AAGAAATATCTACCAAGCCATGCGAGAGCCCATCAAGGAAGTCCCAAGTGCACAAAG 2167
Db 2108 AAGAAATATCTACCAAGCCATGCGAGAGCCCATCAAGGAAGTCCCAAGTGCACAAAG 2167
Oy 1925 ACATGGTCCCTTAAGACCAAGAAATGGGGGTTTACCCAGCTCATGAGGTTTCCAG 1984
Db 2168 ACATGGTCCCTTAAGACCAAGAAATGGGGGTTTACCCAGCTCATGAGGTTTCCAG 2227
Oy 1985 AGTGTGCTCAGCTGTGTGGCTTCTGATCTCGGTGCTGAGGAGCCAGAGGACAGCAAGT 2044
Db 2228 AGTGTGCTCAGCTGTGTGGCTTCTGATCTCGGTGCTGAGGAGCCAGAGGACAGCAAGT 2287
Oy 2045 TGTGTCTCACTTTGTGAGCAACACCCCTGTGTACAGGTTAAAGTTAAAGTTAAACGGGTA 2104
Db 2288 TGTGTCTCACTTTGTGAGCAACACCCCTGTGTACAGGTTAAAGTTAAAGTTAAACGGGTA 2347
Oy 2105 GCGTTCCTCCGAGCATGCTGTGAGATTTGTTGTCATCGCGGATGGAGACACCC 2164
Db 2348 GCGTTCCTCCGAGCATGCTGTGAGATTTGTTGTCATCGCGGATGGAGACACCC 2407
Oy 2165 TGAGGAGATCTTGAGACTGTGATTTTTCAGGGGGCCCCCAAGGCTAGCCAGCCCTCTG 2224
|||||

```

```

Db 2408 TGAGGAGATCTTGAGACTGATGATTTTTCAGGGGGGCCCCCCAGAGGCTAGCAGACCTCTG 2467
Oy 2225 GCGGCTGCAAGGTTAAACAGATCCCTTAACAGAGATGAGACAAACAGCCAGCACCCTG 2284
Db 2468 GCGGCTGCAAGGTTAAACAGATCCCTTAACAGAGATGAGACAAACAGCCAGCACCCTG 2527
Oy 2285 CTGACAGCAGACAGTGGGTCTCAAGAGCTCTGGCCAGACCCCTCCACACCCAGC 2344
Db 2528 CTGACAGCAGACAGTGGGTCTCAAGAGCTCTGGCCAGACCCCTCCACACCCAGC 2587
Oy 2345 CTGCTGTGAAGCAATTCCTGTGACCTGCACACTGTGGCCAGAGAGGCTGTCTGCTACTG 2404
Db 2588 CTGCTGTGAAGCAATTCCTGTGACCTGCACACTGTGGCCAGAGAGGCTGTCTGCTACTG 2647
Oy 2405 TCCGTGAAGAGGGCCCCCAACCGGGGCGGAGTCTTTTAAAGTCAACAGAGTAGTGTCA 2464
Db 2648 TCCGTGAAGAGGGCCCCCAACCGGGGCGGAGTCTTTTAAAGTCAACAGAGTAGTGTCA 2707
Oy 2465 ACTTTCCTGTTGGGAGAGAGAGCCCAATCCGGAGCAGAGAGGGGCTCTGCTTGGCAT 2524
Db 2708 ACTTTCCTGTTGGGAGAGAGAGCCCAATCCGGAGCAGAGAGGGGCTCTGCTTGGCAT 2767
Oy 2525 ATGACCCCTGGGCGGCTCCCTGATGATGCCACACAGGCGCCAGGATTCACCTAGTGGGT 2584
Db 2768 ATGACCCCTGGGCGGCTCCCTGATGATGCCACACAGGCGCCAGGATTCACCTAGTGGGT 2827
Oy 2585 TTGGCAACCTGTGATGTGAGTGTGTAGTGCACATCTGCTTTTGACAGCCAGCCTCCG 2644
Db 2828 TTGGCAACCTGTGATGTGAGTGTGTAGTGCACATCTGCTTTTGACAGCCAGCCTCCG 2887
Oy 2645 TCACACGAGCTGTGAGAAAGATGAGACCAACAAAGGGGGGCGCGTTCCACATGTGCCA 2704
Db 2888 TCACACGAGCTGTGAGAAAGATGAGACCAACAAAGGGGGGCGCGTTCCACATGTGCCA 2947
Oy 2705 AGCCGAGAGAGCAGCAGTGTGCTTTTCCAGTGGGTGCATGAGAACACCGCTCCAGGGA 2764
Db 2948 AGCCGAGAGAGCAGCAGTGTGCTTTTCCAGTGGGTGCATGAGAACACCGCTCCAGGGA 3007
Oy 2765 CTTCCTGAGAGCCCGCTCTGAGACAGAGAGAGAGAAACCTGTGAGTGGAAAGCCAGAA 2824
Db 3008 CTTCCTGAGAGCCCGCTCTGAGACAGAGAGAGAGAAACCTGTGAGTGGAAAGCCAGAA 3067
Oy 2825 GCAAAAGGCGCGGCGGCGAGTCTCTCAGACATGGGGTCCACAGCAAGAAACCCCGGAAT 2884
Db 3068 GCAAAAGGCGCGGCGGCGAGTCTCTCAGACATGGGGTCCACAGCAAGAAACCCCGGAAT 3127
Oy 2885 GCAGCCTTTGCCACACAGCCTGAGACACACCCGTCCTTTTGTCTCTCAGAACAGATGAGCTC 2944
Db 3128 GCAGCCTTTGCCACACAGCCTGAGACACACCCGTCCTTTTGTCTCTCAGAACAGATGAGCTC 3187
Oy 2945 AGGCTAGGCTAGAGAA 2960
Db 3188 AGGCTAGGCTAGAGAA 3203

```

```

RESULT 2
AB006074 3741 bp mRNA linear ROD 18-APR-1998
LOCUS AB006074
DEFINITION Mus musculus mTOP3 mRNA for topoisomerase III, complete cds.
ACCESSION AB006074
VERSION AB006074.1 GI:3061307
KEYWORDS mTOP3; topoisomerase III.
SOURCE Mus musculus (strain: Balb/c) testis cDNA to mRNA.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 3741)
AUTHORS Seki,T.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1997) Takahiko Seki, Tohoku University, Faculty
of Pharmaceutical Sciences, Aoba Aramaki, Aoba-ku, Sendai, Miyagi
980-77, Japan (E-mail: takahima12.pharm.tohoku.ac.jp,
Tel: +81-22-217-6876, Fax: +81-22-217-6873)

```

[illegible]

Db	702	GGAGGAGCTGTGATGGGAAAGCTGAGAACATCGGGCTTTAGATTTACACGATACCAAG	761
Oy	431	CTGTAAAGCCCAATCTCAGAGTGTGGAGACCCGATTTCTGAGATCACACCCCATGCGC	490
Db	762	CTGTAAAAACCAATCTACGGGTGTGAGACCCGGTTCTGTGATCAACGCCACGCCG	821
Oy	491	TCGAGCAGACTCTTGAAAACCTGACCGAGCCTATCAGAGGCTGAGCCATGCTGTGATG	550
Db	822	TCGAAACAGCCTGTGAAAACCTTAACCTAGCCTTACCAAGAGCTGAGTGCAGATGGATG	881
Oy	551	TGAGGCAAGAGCTGGACCTGAGATTGGAGACTGCCTTTACTAGTGTCCAGACCCTGCGGC	610
Db	882	TGAGGCAAGAGCTGGACCTGAGATTGGAGACTGCCTTTACTAGTGTCCAGACCCTGCGGC	941
Oy	611	TTTCAGAGATTTTCTTGAGGTGCTGGACAGACTCATCAGTTACGGCACTCCACT	670
Db	942	TCCAGAGATTTCTCCCTGAGGTGCTGGACAGAGCTCATCAGCTACGGCACTTCCCACT	1001
Oy	671	TCCCAACCTGGGCTTTGTGTGAGGGGTTCAACCACTTACAGGCTTTTGACCAAAA	730
Db	1002	TCCCAACCTGGGCTTTGTGTGAGGGGTTCAACCACTTACAGGCTTTTGACCAAAA	1061
Oy	731	TCTTCCACAGATTAAAGTAACTCATGACACCAAAAGATGGTATCGTAGATTCAACTGGA	790
Db	1062	TCTTCCACAAATTAAGTAACTCATGACACCAAAAGATGGGACCGTAGATTCAACTGGA	1121
Oy	791	AAAGGCAATGCACTTTTAACCAACAGGCGTGGCTAGTCTCATGAGTTGTGTGAGAG	850
Db	1122	AAAGGCAATGCACTTTTAACCAACAGGCGTGGCTAGTCTCATGAGTTGTGTGAGAG	1181
Oy	851	ATCCCATGGCAACTGTGTGAGAGTCAAGTCTAAGCCCAAGAGCAATGGCGCCTCAAG	910
Db	1182	ATCCCATGGCGACTGTGTGAGAGTCAAGTCTAAGCCCAAGAGCAATGGGACGCGCAGG	1241
Oy	911	CCTTGAGACACTGTGAGACTTGAGAACGTGGCTTCTCGAAGTTGGAGATTAATGTAAAG	970
Db	1242	CTTTGGACACGTGTGACTGTAGAAACCTGGCATCTCTCGAAATTTGGAATTAATGTCAAG	1301
Oy	971	AAACCATGAGAGTGTGAGAACCTCTACACTCAAGGTTACATCAGCTATCCCCCAACAG	1030
Db	1302	AAACCATGAGAGTGTGAGAACCTCTATNCAACAGGTTACATCAGCTATCCCCCGGACAG	1361
Oy	1031	AAACAAACATTTTCCAGAGACTTAAACCTGACGCTGTGGTGTGAACAGCAGACCCCGC	1090
Db	1362	AAACAAACATTTTCCCAAAAGACTTAAACCTGTGTTGATGTGAACAGCAGACCGTGG	1421
Oy	1091	ATTCACGCTGGGGGCGCTTGGCCACAGCACTTGTAGAGGGGGTGTGCCACCCACCA	1150
Db	1422	ATTCACACTGGGGGGCTTTTGTCTCAGACCACTTGTAGAGGAGAGTGGCCTTACTCCCGAA	1481
Oy	1151	ATGGAACAAAGTCTGACCAAGCTACACCTTCCATTCCACCCACCAAAATACACCAACA	1210
Db	1482	ACGGGAGCAAGTCTGATCAGACTCACCTTCCCTCACCCCAAAATACACCAACGAGCT	1541
Oy	1211	TACAGGGAATGAACACCGCACTGACGAGTTATTTGTGGCAATTTCTGGCTGTGCT	1270
Db	1542	TGCGAGGAGATGATCGGCGCACTGACGAGTTCAATTTTGCCATTTCTCGGCTTGCTGCT	1601
Oy	1271	CCCGAGATGCTCAGGGGCGAGAGCAACCAAGTGGAGATGCATCCTCAGAGACGCTTGG	1330
Db	1602	CTCAGGAGCGTCAAGGGGCAAGAGACTAACGGTAGATTCAGATTTGCCAAGAGCGCTTGG	1661
Oy	1331	TGGCCCATGCTCATGATTTCTGGCCCGAAACTATCTGGATGTGATCATATGATCACT	1390
Db	1662	TGGCCCATGCTCATGATTTCTGGCCCGAAACTATCTGGATGTGATCATGATGATCACT	1721
Oy	1391	GGAGTGAACAAGATCTCCCTGTCTATGAGCAAGGATCCCACTTTGACGCCACAGACCTGG	1450
Db	1722	GGAGTGAACAAGCTCTCCCGTCTATGAAACAAAGCTCCCACTTCCAGGCCACAGACTGGC	1781
Oy	1451	AGATGTGAGACGGGAGACCAAGCCCAACCAAGCTGCTACCGAGGCGCACTCATTTGCC	1510

QY	2585	TTGGCAACCCCTGATGATGAGCAGTGGTAATGGCAGATCTCGCTTGTGAGCCAGCCCTTCGG	2644
Db	2992	TTGGCAGCCCTTGGCAGCGACAGTGTATGAGAGGTACACCCCTGCTGTGGGGGCGAGCTGTCTG	2981
QY	2645	TCACACGCACTGTGCAGAGAATGGACCCCAACAAGGGGCGCCAGTCTCCACATGTGCCA	2704
Db	2982	TCACACGCACTGTTCAGAAAGATGGACCCCAACAAGGACGCCAGTCTCCACACCTGTGCCA	3041
QY	2705	AGCCGAGAGACAGCAGTGTGGCTTTTTCAGTGGGTGATGAAACACCCGCTCCAGGGA	2764
Db	3042	AGCCACGAGAGCAGAGTGTGGCTTTTTCAGTGGGTGATGAAATGTGGCCCCAGGAA	3101
QY	2765	CTTCTGAGACCCCGCTCTCGACAGAGACAGAGAAGAACCCCTGAGTCCGAGCCAGAA	2834
Db	3102	GTTTTCGACGCCCTGTCATATGGCCAGAGGACAGAGAAAGCCCAAGACCAAGAGCTGCMA	3161
QY	2835	GCAAAAGCCCGCCGGCCAGTCTCTCAGACATGGGGTCCACAGCAAGAAACCCCGAAAT	2884
Db	3162	GCAAAAGACCAAGAGCTGTCTCTCAGATGCAGGGTCCACAGTAAAAAGCCCTCGAAAT	3221
QY	2885	GCAGCCTTGGCAGCAGCCTGGACACACCCGCTTGTTCGTCACAAACGATGAGCTC	2944
Db	3222	GTAGCCTTGTTCACAGCCTGGACACACCCGACCTTTGTCTCAGAAACGATGATAGC	3281
QY	2945	AGCGTAGGCTAGAGAA	2960
Db	3282	CGGGAGAGGTGGAAGAA	3297
RESULT 3			
LOCUS	HSF10178/c	132486 bp	DNA linear PRI 17-JUN-2001
DEFINITION	Homo sapiens chromosome 17 sequence from PAC 178f10 region		
ACCESSION	D17S842-D17S953 map 17p11.2, complete sequence.		
VERSION	AL035367		
KEYWORDS	AL035367.5 GI:13752116		
SOURCE	HTG.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	1 (bases 1 to 132486)		
REFERENCE	Ramser,J., Mueller,I., Sudbrak,R., Kosura,A., Radejof,U., Hennig,S., Francis,F., Steffens,C., Klein,M., Seraniki,P., Poustka,A., Lehrach,H. and Reinhardt,R.		
AUTHORS	Unpublished		
JOURNAL	2 (bases 1 to 132486)		
REFERENCE	Submitted		
AUTHORS	Submitted (05-FEB-1999) MPIMG, Abt Lehrach, Max Planck Institut		
TITLE	Fuer Molekulare Genetik, Innestrasse 73, Berlin, 14195 Germany		
JOURNAL	On Apr 22, 2001 this sequence version replaced gi:1921151.		
COMMENT	Clone received from the Resource Centre of the Human Genome Project at the Max-Planck-Institute for Molecular Genetics		
FEATURES	1..132486.		
Source	Location/Qualifiers		
	1..132486		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="17"		
	/map="17p11.2"		
	/clone="PAC RPCI-1.178f10"		
	/clone_lib="RPCI1,3-5 Human PAC library, originating		
	Institute: Roswell Park Cancer Institute, creator: Pieter		
	de Jong, P. Ioannou"		
	/note="region between markers D17S842-D17S953"		
	1		
	/note="SP6_end:PAC RPCI-1 178f10"		
	132486		
	/note="T7_end:PAC RPCI-1 178f10"		
BASE COUNT	31555 a 34946 c 34918 g 13067 t		
ORIGIN			

Query Match 23.1%; Score 686; DB 9; Length 132486;
Best Local Similarity 100.0%; Pred. No. 6,6e-172;
Matches 686; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2077 AGGTTAAAGTTAAAGTTAAAGCGGGGAGCTTCCCGGACCAATGCTCTGGAGTTGTT 2136
|||||
Db 53784 AGGTTAAAGTTAAAGTTAAAGCGGGGAGCTTCCCGGACCAATGCTCTGGAGTTGTT 53725

QY 2137 TGGTCATCGCGCGATGCGAGACACCCCTGAGGAGATCTCGACCTGAGATTTTCAGGG 2196
|||||
Db 53724 TGGTCATCGCGCGATGCGAGACACCCCTGAGGAGATCTCGACCTGAGATTTTCAGGG 53665

QY 2197 GGGCCCCCAGGCGCTAGCCAGCCCTGCGCCGCTGAGAGGTTAAACAGTCCCTGACAG 2236
|||||
Db 53664 GGGCCCCCAGGCGCTAGCCAGCCCTGCGCCGCTGAGAGGTTAAACAGTCCCTGACAG 53605

QY 2257 ATGACAAAGCCAGACACCCCGACCTGCTGACAGCAGACAGAGTGGGTCTCTAAAGCT 2316
|||||
Db 53604 ATGACAAAGCCAGACACCCCGACCTGCTGACAGCAGACAGTGGGTCTCTAAAGCT 53545

QY 2317 CTGGCCCGAGACCTCCACACACCCAGCGCTGCTGTAAGCAATTTCTGACCTGCAAC 2376
|||||
Db 53544 CTGGCCCGAGACCTCCACACACCCAGCGCTGCTGTAAGCAATTTCTGACCTGCAAC 53485

QY 2377 TGTGGCCAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2436
|||||
Db 53484 TGTGGCCAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53425

QY 2437 TTTCTTTAAGTGAAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2496
|||||
Db 53424 TTTCTTTAAGTGAAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 53365

QY 2497 GGAACAGAGAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2556
|||||
Db 53364 GGAACAGAGAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53305

QY 2557 CCAAGGCCAGAGGATCCACATAGTGGGTTTGGCAACCTGCTGCTGCTGCTGCTGCTG 2616
|||||
Db 53304 CCAAGGCCAGAGGATCCACATAGTGGGTTTGGCAACCTGCTGCTGCTGCTGCTG 53245

QY 2617 ACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2676
|||||
Db 53244 ACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53185

QY 2677 AAGGGCGCCAGTTCACACATGTGCCAAGCGAGAGAGCAGCAGTGTGGCTTTTCCAG 2736
|||||
Db 53184 AAGGGCGCCAGTTCACACATGTGCCAAGCGAGAGAGCAGCAGTGTGGCTTTTCCAG 53125

QY 2737 TGGGTGATGAGAAACACCGCTCCAGG 2762
|||||
Db 53124 TGGGTGATGAGAAACACCGCTCCAGG 53099

RESULT 4
AF255733 3891 bp mRNA linear INV 21-MAY-2000
LOCUS Drosophila melanogaster DNA topoisomerase III alpha (CG10123) mRNA,
DEFINITION complete cds.
ACCESSION AF255733
VERSION AF255733.1 GI:7960301
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Plank,J.L., Reineke,J.C., Wilson,T.M. and Hsieh,T.-S.
TITLE Drosophila melanogaster topoisomerase III alpha
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3891)
AUTHORS Plank,J.L., Reineke,J.C., Wilson,T.M. and Hsieh,T.-S.

TITLE Direct Submission
JOURNAL Submitted (13-APR-2000) Biochemistry, Duke University Medical
Center, Research Drive, Durham, NC 27710, USA
FEATURES
Source
1.3891
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2"
/map="2L; 37E3"
/dev_stage="embryo"
1.3891
/gene="CG10123"
/note="top3a"
37.3789
/gene="CG10123"
/codon_start=1
/product="DNA topoisomerase III alpha"
/protein_id="AAE71288.1"
/db_xref="GI:7960302"
/translation="MKAVLTCFPPVYASHRRYCANPQAMKLYLNAEKNDAKTIAGL
LSMAQOREGYSVYKVFDEAPVVGONAKMYTSVSGHMOALAOVSYKNNRTVP
RSLFDPVREKVSQSDYPIKRTLEREVGCGGLITWDCRGENIGYEITIDCAIX
PNISVYRATSEITTVAVRRALQOLQOPKROSDAVDRFELDLPTGAATIRQTRRL
QRFPKRIADKLISYSCQIPRTGPAVERKETEAVRSSEPFKIKYLIHTLDLTFEEN
WARKNLEKACENYLLCLAEPPRALVESVTVKRRKRRPTPLDTEVERKLSKRL
KLSAKETMTIAERLYTKGFTSYPRTEINQSKFALAPLVEMOTGHRDWAERQVIE
WCPNPRGNKSDAHPHPTKLAENLOGNEARVIELVNHFTLACYSKAVGSETLVH
IDLAGEKFTANGLVIERNYLDVYVYDGNANAKOIHHEENQOREPVEVLSHEGATAP
PLTEADLILMEKHGIGDPAHAENHNTKEGIVGLDKRGFLPVGIVGMLYEGED
AMELALAKPOLRAEFELDKLIGOGKRDGIVFTFQIARUKAOYQITDKITAMDKI
SARINTPRAANSVQSDADSSASHGIIQSPICPPRNCNAPRLARKKNDQYICGN
NFPDCRNAWVLPTECKDASVLDCECTPGGIGTGMFLFRSTPIYKGFPGSWITIC
LPQDNLFRFTFNINLDSVYKIVGIVGVEVRGGGPGPBGGGSGGAGSGGSSGPG
GGGGGGGGSGGSGTGG
KPKRTKEPRAAPKRTSSKSSGSIRESFTSAAPTNSAGSLDEFSDNGFEDAMLEA
AESVSSSOPKRTISVPLDDDDIAAFAADDAEFELVNGTPTESNDQDOLKSL
EWIKEDKDERPMLMCTERRASIGTAAPPPKPAKRRPMSVSRSDPSVSES
EYVLTGCGQOPARQNTVRKNGPMLGLYIKCPRPDCNFFQNADEPPSSAKSKNSG
APQSTTSWDSNRVYTLPSIQOSNORQSSMRNSSSVYITQTKQDERNATPBG
GERVMNCQQLAQLTVRRDGMQGRPFACPTREKSGFFKMGDNOGASSTWSG
SANRNPGRSOPPAITSIDGPKTRRCGLCKRKHGTRNKKCPRKDEFDM"

BASE COUNT 1030 a 973 c 1095 g 793 t

ORIGIN

Query Match 20.0%; Score 594.6; DB 3; Length 3891;
Best Local Similarity 59.6%; Pred. No. 1.1e-147;
Matches 1043; Conservative 0; Mismatches 699; Indels 9; Gaps 2;

QY 36 GAAAGTCTCTGTGTGGCCGCAAAAACGACGCGCCCAAGGGATGCGGACCTGCTGTC 95
|||||
Db 114 GAAAGTCTCTGTGTGGCCGCAAAAACGACGCGCCCAAGGGATGCGGACCTGCTGTC 173

QY 96 AAACGATCGCATGAGCGGAGAGAGACATTTCAAAATTCACAAGAATCTATGAATTTGA 155
|||||
Db 174 AAATGCGCGCAGCACAAGCGCGGAGAGCTACGCTATACACAAGGCTCTTCGATTGCA 233

QY 156 TTATCATCTGTATGGCCAGAAATGTTTACATGTAATGATGATTCAGTTCCTGACATTTACT 215
|||||
Db 234 GGGCCCTGTGGCGGCTGAGAAATGTTATGACCTCCGTATCCGGCCACATGAT 293

QY 216 GGCATCATGATTTCCAGATGCAAGTTTCGAAAATGGCAAGCTGCACACCTCTTGCTCTT 275
|||||
Db 294 GCAAGTGGCATTTCCAGGTTGCTTCAAAAATTTGGAGAAGGTGGATCCGGCTCTTATTT 353

QY 276 TGAAGCAGAAATTTGAAGTACTGCCAGAAATTTTGTAGACATCAAGAAACTTTTGA 335
|||||
Db 354 CGACGCGCGCGGTGAAAAGGCGGTGGGCTGGATTTATGACCTTATCAAAAGACCTTTGGA 413

QY 336 ACGAGAGACTCGCCAGTGCAGGCTGTGATCTGGACCTGACTGTGATAGAGAAGCGGA 395
|||||
Db 414 ACGGAGAGTGGCGGGCTGTGACGAGGATTTATCTGACGAGATTTGCATCGCGAGGCGGA 473

QY 396 AAACATCGGGGTTTGAGATTTATCCACGCTGTAAAGCTGTAAAGCCCAATCTGCAGGTTGTT 455

```

Db 474 GAACATTTGGCTAGAGATTATGATGTGTGCGCTATCAAAACCAATATTTGCGTTTA 533
Qy 456 GCGAGCCCATTTCTCTGAGATCATACACCCATCCGCTCAGACAGCTTTGAAAACCTGAC 515
Db 534 TCGTGCACATTTCTCCGAGTTTACCAACGGTGGTGGCTCGCGCTCTGAGCAATTTGGG 593
Qy 516 CGAGCCTGATCAGAGGAGGTGAGCTGTGATGTGAGGACGACCTGGAGCTGAGAGAT 575
Db 594 GCAGCCCGAAGAAAGCAGAGCGATGAGTGTGCGACGGAAGCTGGATTTGCGAAC 653
Qy 576 TGGAGCTGCTTACTAGATTCAGACCCGCGCTTCAGAGAGATTTCTGAGGTGCT 635
Db 654 TGGTGGCCGATACCCCGGTTTACAGAAATGCCCTACAAAGCTGTGTTCCCGAANAAGAT 713
Qy 636 GCGAGACAGCTCATGATGATGAGGACGTGCGACGTTCCGACAGTGGCTTGTGTGGA 695
Db 714 AGCTGATTAAGCTCATATTCATGGAAGCTGCGCAAAATTCACCTGGATTCGCGCGA 773
Qy 696 GCGGTTCAAAAGCCATTGAGGCTTTTGTACAGAAATCTTCCACAGAAATTAAGTAATCA 755
Db 774 AGCGTACAAAGAGATGAGGCGCTTGTATCCGAGCCATTTTGGAAAGATTAAGTTCTGCA 833
Qy 756 TGACACAAAGATGTATGCTAGATTTCAACTGAAAAGGATCGACTCTTTAAACCAAC 815
Db 834 CACAAATGATGACTTACGAGGTGAGATTTAATTTGGCCCGAAGCCGACTCTTCATTAAGGA 893
Qy 816 GCGTTGCTAGTTCTCTATCAGTTGTGTGTG-----AGATCCAGTGGCAACTGTGTG 869
Db 894 AGCTGTGAAAACATATTTGCTGCTGCGCGTGGCGAACCAGATCTAGAGCGGTGCGGA 953
Qy 870 AGAGTCAGATCTAAGCCCAAGAGCAAGTGGCGGCTCAAGCTTGTGACACTGTGAGCT 929
Db 954 GAGCGTACCGCTTAACCCCAACAAATGCGGCCCAACCTCCGCTGATACCTGTGAAA 1013
Qy 930 TGAGAGCTGGCTTCTCGAAGTTGAAATTAATGCTAAAGAAACATGAGATTTGCTGA 989
Db 1014 GGAAGAGCTTGGCTCGAGAAACTCAAGCATATGCGCAAGAGAACATGACATAGCCCA 1073
Qy 990 GAAGCTCTACACTCAAGGGTACATCAGTATCCCGCAAGCAAAACAATTTTCCAG 1049
Db 1074 AAATATATATACCAAGGGTTTCACTAGCTATCCCGCTACGGAACCAACAGATTTCTCA 1133
Qy 1050 AGACTTAACCTGAGCGGTGTGTGTGAAACAGACAGCCCCGATCCAGCGTGGGGGCGCTT 1109
Db 1134 GAGGTTGCTGTGGACCACTAGTTAGATGAAAGGGGTACCGCGACTGGGGGAGCATTT 1193
Qy 1110 TGGCCAGAGCATTTCTAGAGGGGGGTGTCGCCACCCAGCAATGAGGAACAAGTCTGACA 1169
Db 1194 TGGCCAGCGGGTATGATGAGTG---GGACCAAAATCTCTCGAAAGGTTAACAAATCGGATCA 1250
Qy 1170 AGCTCAACCTCCCATTCACCCCAACAATACACCAACACTTACAGGAGATGAGACGC 1229
Db 1251 GGGCCATCTCCCAATTCATCCCAACAGCTGGCAGAAACCTCAGGGAACGAGGCGCG 1310
Qy 1230 ACTGTACAGATTATTTGTTGGCCATTTCTGCGTTGCTGCTGCCAGATGCTCAGGGGCA 1289
Db 1311 AGGTACAGAACTGTGTCTCGGACCTTCTCGCTGCTGTAGAGAAAGATGCACTGCGCTC 1370
Qy 1290 GGAAGACACAGTGGAGATGACATCGCTCAGAAAGCTTTGTGGCCATGGCTCATGAT 1349
Db 1371 CGAAACCTTTGTCATATGACATTTGCGGGTGGAAAGTTACGCGCAATGCTGTGTTAT 1430
Qy 1350 TCTGGCCGCAAACTATCTGAGATGTGATCCATATGATCACTGAGTGAAGATCTCTCC 1409
Db 1431 TCACGACGAACAACTACTAGATGTTATGTTATGACAACTGAGTGGCCAGAGATGACA 1490
Qy 1410 TGTCTATGAGCAAGATCCCACTTTCAGCCGACACCGTGGAGATGTTGAGCGGGAGAC 1469
Db 1491 TCACATATGAAAAGCGGAGGTTTGGAGCCACGGAAGTGTGCTCCACCAAGTGTCCAC 1550
Qy 1470 CACCCACCAACAGCTGTCACGAGCGGACGACCTATTTGCCATGGAAGAGATGGCAT 1529

```

```

Db 1551 CACTGCCCGCCCGCTGTCAGCGAGCGGATTTGATTGCACTGATGGAAGAGATGGAT 1610
Qy 1530 TGGTACGATGCGCACTCATGCGAGACATCGAGACCATCAAGCCCGATGATGCTGG 1589
Db 1611 TGGTACGATGCGCACGACGCTGAGACATTAACAGATCAAGGAAGCGCGATCATCGG 1670
Qy 1590 CCTCACCCAGACAGCGGTTCTCCCTGGGACGACCTGGGAGGAGCTGTGGAAGTTA 1649
Db 1671 AGTGTGATGATAGGCGCTTTTGTGCTACCTGAGATTAATGGAATGCACTGATAGGATA 1730
Qy 1650 TGATTCATAGGCGCTATGAATGCTAAGCCTGACCTCCGCGCTGAGATGGAAGCTGATC 1709
Db 1731 CGATGCGATGAGACCTGCGCTTGCACAGCACAACTCTGCTGAGATTTGAATTTGATC 1790
Qy 1710 GAAGCTGATCTGTGATGCGCAAAAGACAAATTTGTGTTCTAAGCAGCACTGACAA 1769
Db 1791 GAACATTAATTTGGCAGGCGCAGAAAGATCCCAAGAGTGGTCTGACCGAACATGACCA 1850
Qy 1770 ATACAGCAGG 1780
Db 1851 GTACAAAGCAGG 1861

RESULT 5
AC097545/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-68E7, *** SEQUENCING IN PROGRESS ***
ACCESSION
AC097545
VERSION
AC097545.3 GI:11973505
KEYWORDS
HTG: HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 88718)
REFERENCE
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsdbrook,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbista,J.,
Benton,J., Blinze,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gotrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Bulky,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J.,
Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mathlony,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,S., Nguyen,N., Nickerson,E., Nwokwenwo,S.,
Oguth,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,N.,
Stone,H., Sutton,A., Svatik,A., Tabori,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telifrod,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vanson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

```


TITLE
JOURNAL
AUTHORS
REFERENCE
TITLE
JOURNAL
COMMENT

Watlington, S., Williams, G., Williamson, A., Wlarczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (phases 1 to 88718)
Worley, K.C.
Direct Submission
Submitted (19-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064400.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHY0
Center clone name: CH230-68E7
----- Summary Statistics
findhrp1list
Assembly program: Phrap; version 0.990329first call to
Consensus quality: 74659 bases at least Q40
Consensus quality: 79368 bases at least Q30
Estimated insert size: 66722; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 0.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4658: contig of 4658 bp in length
* 4659 4758: gap of unknown length
* 7911 7910: contig of 3152 bp in length
* 8011 8010: gap of unknown length
* 11829 11828: contig of 3818 bp in length
* 11929 11928: gap of unknown length
* 15314 15313: contig of 3385 bp in length
* 15414 15413: gap of unknown length
* 18540 18539: contig of 3126 bp in length
* 18640 18638: gap of unknown length
* 21283 21282: contig of 2643 bp in length
* 21383 21382: gap of unknown length
* 24341 24340: contig of 2958 bp in length
* 24441 24440: gap of unknown length
* 26397 26396: contig of 1956 bp in length
* 26497 26496: gap of unknown length
* 29011 29010: contig of 2514 bp in length
* 30511 30510: contig of 1400 bp in length
* 30611 30610: gap of unknown length
* 32920 32919: contig of 2309 bp in length
* 33020 33019: gap of unknown length
* 34735 34734: contig of 1716 bp in length
* 34835 34834: gap of unknown length
* 36525 36524: contig of 1690 bp in length
* 36625 36624: gap of unknown length
* 38856 38855: contig of 2231 bp in length
* 38956 38955: gap of unknown length
* 38957 38956: gap of unknown length
* 40235 40234: contig of 1279 bp in length
* 40336 40335: gap of unknown length
* 41823 41822: contig of 1488 bp in length
* 41824 41823: gap of unknown length
* 41924 41923: gap of unknown length
* 43349 43348: contig of 1426 bp in length

43350 43449: gap of unknown length
* 43450 45217: contig of 1768 bp in length
* 45218 45317: gap of unknown length
* 45318 47327: contig of 2010 bp in length
* 47328 47427: gap of unknown length
* 47428 49395: contig of 1968 bp in length
* 49396 49495: gap of unknown length
* 49496 51028: contig of 1533 bp in length
* 51029 51128: gap of unknown length
* 51129 52413: contig of 1285 bp in length
* 52414 52513: gap of unknown length
* 52514 54378: contig of 1865 bp in length
* 54379 54478: gap of unknown length
* 54479 55757: contig of 1279 bp in length
* 55758 55857: gap of unknown length
* 55858 57804: contig of 1947 bp in length
* 57805 57904: gap of unknown length
* 57905 59300: contig of 1396 bp in length
* 59301 59400: gap of unknown length
* 59401 60588: contig of 1188 bp in length
* 60589 60688: gap of unknown length
* 60689 62433: contig of 1745 bp in length
* 62434 62533: gap of unknown length
* 62534 64174: contig of 1641 bp in length
* 64175 64274: gap of unknown length
* 64275 65286: contig of 1012 bp in length
* 65287 65386: gap of unknown length
* 65387 67298: contig of 1912 bp in length
* 67299 67398: gap of unknown length
* 67399 68966: contig of 1568 bp in length
* 68967 69066: gap of unknown length
* 69067 70250: contig of 1184 bp in length
* 70251 70350: gap of unknown length
* 70351 71676: contig of 1326 bp in length
* 71677 71776: gap of unknown length
* 71777 72899: contig of 1123 bp in length
* 72900 72999: gap of unknown length
* 73000 74690: contig of 1691 bp in length
* 74691 74790: gap of unknown length
* 74791 75850: contig of 1060 bp in length
* 75851 75950: gap of unknown length
* 75951 77055: contig of 1105 bp in length
* 77056 77155: gap of unknown length
* 77156 78265: contig of 1110 bp in length
* 78266 78365: gap of unknown length
* 78366 79576: contig of 1211 bp in length
* 79577 79676: gap of unknown length
* 79677 81067: contig of 1391 bp in length
* 81068 81167: gap of unknown length
* 81168 82201: contig of 1034 bp in length
* 82202 82301: gap of unknown length
* 82302 83371: contig of 1070 bp in length
* 83372 83471: gap of unknown length
* 83472 84840: contig of 1369 bp in length
* 84841 84940: gap of unknown length
* 84941 86137: contig of 1197 bp in length
* 86138 86237: gap of unknown length
* 86238 87452: contig of 1215 bp in length
* 87453 87552: gap of unknown length
* 87553 88718: contig of 1166 bp in length.

FEATURES
source
1..88718
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-68E7"

BASE COUNT 21517 a 20394 c 20073 g 22076 t 4658 others
ORIGIN
Query Match 15.08; Score 446; DB 2; Length 88718;
Best Local Similarity 78.0%; Pred. No. 1e-107;
Matches 552; Conservative 0; Mismatches 150; Indels 6; Gaps 1;
OY 2077 AGCTTAAGTTAAGTTAAGCGGCGTACGCTTCCCGGACCAATGCGAGCTTGT 2136

DB 144126 GGAGGTACACCTTCTTGGGCGAGCTGTCTACACGAGCTTCAGAGATGGA 144185
QY 2671 CCCAACAAAGGCGCCGACATTCACACATGTGCGCAAGCCGAGACAGACAGTGTGCTTT 2730
DB 144186 CCCAACAAAGGCGACATTCACACCTGTGCGCAAGCCGAGACAGACAGTGTGCTTC 144245
QY 2731 TTCCAGTGGTGATGAGACACCGCTCCAGG 2762
DB 144246 TTTCAGTGGTGATGAGACATGTGCCCCAGG 144277
RESULT 7
AC069534/c 213913 bp DNA linear HTG 16-OCT-2001
LOCUS Mus musculus chromosome 11 clone RP23-5201, WORKING DRAFT SEQUENCE,
DEFINITION 5 unordered pieces.
AC069534
AC069534 AC069534 20 GI:14547767
AC069534
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 213913)
AUTHORS Metzger,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okunou,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Ranney,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,Z., Liu,W., Louised,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mel,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Ogih,M., Parish,B.,
Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Wellington,S.,
Williamson,A., Wrenford,G., Zhou,X., Bouck,U., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 213913)
AUTHORS Worley,K.C.
JOURNAL Direct Submission
COMMENT Submitted (02-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:13162471.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MAFN
Center clone name: RP23-5201
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 51% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 215914 bases at least Q40
Consensus quality: 218176 bases at least Q30
Consensus quality: 219553 bases at least Q20
Estimated insert size: 215987; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agnos-fp estimation
Quality coverage: 8.2x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
80577: contig of 80577 bp in length
80578 80677: gap of unknown length
80678 157388: contig of 76711 bp in length
157389 157488: gap of unknown length
157489 192191: contig of 34703 bp in length
192192 192291: gap of unknown length
192292 206716: contig of 14425 bp in length
206717 206816: gap of unknown length
206817 213913: contig of 7097 bp in length.
Location/Qualifiers
1. 213913
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-5201"
BASE COUNT 58068 a 51330 c 50253 g 53853 t 409 others
ORIGIN
Query Match 14.6%; Score 434.8; DB 2; Length 213913;
Best Local Similarity 77.9%; Pred. No. 1.1e-104;
Matches 539; Conservative 0; Mismatches 147; Indels 6; Gaps 1;
QY 2077 AGGTTAAAGTTAAATTTAAAGCGGCGTGTGCTTCCCGACCATGCTCTGAGTTTGT 2136
DB 88393 AGGTTAAAGTTAAATTTAAAGCGGCGTGTGCTTCCCGACCATGCTCTGAGTTTGT 88394
QY 2137 TGTCTCATCGGCGGATGCGAGCAGACCCCTGAGGAGATCTTGACCTGATTTTCAGG 2196
DB 88333 GAGTTCATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 88274
QY 2197 GAGCCCCCAGGCGCTAGGCGAGCCCTCTGCGCGCTGCGAGGCTTACCATGCTTCC 2256
DB 88273 GCTTCCCAAGAGCTAGGCGAGCCCTCTGCGCGCTGCGAGGCTTACCATGCTTCC 88214
QY 2257 ATGACACAGCAGCA-----CCCCAGCCTGTGTGACAGCAGCAGCTGGCTCTCA 2310
DB 88213 ATGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 88154
QY 2311 AAGGCTTGGCCCGACAGCCCTCCACCACCCAGCGCTGTGTGAAAGCAATTTGTGACC 2370
DB 88153 AAGGCTTGGCCCGACAGCCCTCCACCACCCAGCGCTGTGTGAAAGCAATTTGTGACC 88094
QY 2371 TGCACCTGTGCGCGAGGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2430
DB 88093 TGCACCTGTGCGCGAGGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88034
QY 2431 CGGACATCTTAAATGAGAGGAGGATGAGCAGCAGTCTTCTGTTGGGAGAGAGAGCC 2490
DB 88034 CGGACATCTTAAATGAGAGGAGGATGAGCAGCAGTCTTCTGTTGGGAGAGAGAGCC 87974
QY 2491 AATCCGGAGAGAGAGGCTCTGCTTGGCATATTAAGCCCTGGGCGCTCTCTGGA 2550
DB 87973 CATTCACATGAG 87914
QY 2551 TGGCCACAGGCGCAGGAGATCACTAGTGGTGGTGGCAACCCGATGATGAGAGAGT 2610
DB 87913 TGGCCACAGGCGCAGGAGATCACTAGTGGTGGTGGCAACCCGATGATGAGAGAGT 87854
QY 2611 AGTGACATCTGCTTGGAGCGAGCCCTGCTGACAGAGAGAGAGAGAGAGAGAGAG 2670
DB 87853 GAGGTACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 87794
QY 2671 CCCAACAAAGGCGCCGACATTCACACATGTGCGCAAGCCGAGACAGACAGTGTGCTTT 2730
DB 87793 CCCAACAAAGGCGCCGACATTCACACATGTGCGCAAGCCGAGACAGACAGTGTGCTTT 87734
QY 2731 TTCCAGTGGTGATGAGACACCGCTCCAGG 2762

```

Db 87733 TTTGAGTGGTGATGAGAAATGTCGCCACAG 87702
|||||
RESULT 8
AC005428/c 87835 bp DNA linear INV 19-OCT-1998
LOCUS
DEFINITION Drosophila melanogaster DNA sequence (PI DS05231 (D342)), complete
sequence.
ACCESSION AC005428
VERSION AC005428.1 GI:3769307
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (Subclones in sac from P1 clone DS05231
(D342)) DNA.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 87835)
Celisner,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasli,S.R., Katta,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,E.,
Snir,E., Twomey,B., Wan,K.H., Zhang,R., Zieran,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 2R, region 37E1-37E2
Unpublished (1997)
2 (bases 1 to 87835)
Celisner,S.E., George,R.A., Galle,R., Swirskas,R.R., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasli,S.R., Katta,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,E.,
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Oct 19, 1998 this sequence version replaced gi:3764040.
COMMENT
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://fruitfly.berkeley.edu/sequence/) or send
email to bdg@fruitfly.berkeley.edu.
Library location: 4f-55.
Location: Qualifiers
1. 87835
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="37E1-37E2"
/clone="PI DS05231 (D342)"
BASE COUNT 23498 a 20241 c 20288 g 23808 t
ORIGIN
Query Match 14.6%; Score 433.4; DB 3; Length 87835;
Best Local Similarity 55.7%; Pred. No. 2.4e-104;
Matches 995; Conservative 0; Mismatches 671; Indels 119; Gaps 4;
112 CGGAGAGAGGACTTTCACAAATTCACAAAGACTCTTGAATTTGATTATCATTGATGATGC 171
Db 30898 CAGCCGAGAGGCTACTCGGTGACACAAAGGCTTCGATTTGAGAGCCCTTGCGCGGT 30839
Oy 172 CAGATGTATACACTGATGATGACTTCAGTTTGTGACATTTACTAGCTCATGATTTCCAG 231
Db 30838 CAGATGTCCAAATATGTTATGACCTCCGATTCGCCGACATATGACAGTGCATTTCCAG 30779
Oy 232 ATGCAATTTGCAAAATGCGAGAGCTGCAACCCCTTTGTCCTTTTGAAACAGAAATTTGAA 291

```

Db	30778	GTGCTTACAAAATTTGGAGAAAGGTGGATCCGCGCTTTTATTTGACCGCGCGGAGAA	30719
QY	292	AGTACTCGCCAGAGAAATTTTGTAGACATCAAGAAAACCTTTGGAACGAGACACTGCGCAG	351
Db	30718	AAGGGCGTGGCGTCGGATTATGAGCCCTATCAAAAGAGACCTTGGMAACGAGTGGCGGCG	30655
QY	352	TCCCAGAGCCTCTGTGATCTGAGCTAGCTGTGATAGAGAAAGCGGAAACATCGGGTTTGAG	411
Db	30658	TGTCAGAGGATTGATTATCTGAGACGATTGCGATTCGGAGGCGGAGAACTTGGCTACGAG	30599
QY	412	ATTATTCACGTTGTGAAGCTGTAAAGCCCAATCTGCAGGTGTGCGAGGCCGATTCTCT	471
Db	30598	ATTATTGATGTGTGTCGCCCTATCAAAACCGAATATTTCGGTTATCTGTCGCCACTTTCCTG	30538
QY	472	GAGATCACACCCCATTCGCCGTGAGAGACGTTGTGAAAACCTGACCGAGCCTATGAGAG	531
Db	30538	GAGATTACACGCGTGGCGGTGGCTGGGCTCTGACAGCAATTTGGGACGCGGACAAAAG	30479
QY	532	GTGAGCGATGCTGTGGATGTGAGGACGAGAGCTGAGACCTGAGGATTTGAGACTCTTACT	591
Db	30478	CAGAGGATGACAGTGGATGTGGCGACGGAACATGGAATTTGCAACTGTGTGCGCGCATCAC	30419
QY	592	AGGTTCCAGACCCCTCGGGCTTCAGAGGATTTTTCCTGAGGTCGTGGCAGACAGCTCATC	651
Db	30418	CGGTTTCAACAATGGCGCTCAAGCTGTCGTTCCCGGAAAATATCTATTAAGCTCAT	30355
QY	652	AGTTACGGAGAGCTGCCAGTGTCCCACTGACACGCTTTGTGGTGGAGCGTTCAAGACCT	711
Db	30358	TCCTATGGAAGCTGCCCAATTTCCACACCTCGGGATTGTCGCCGACGGTACAAAGGATT	30299
QY	712	CAGGCTTTTGTACACAGAAATCTTCCACAGAAATTA-----	746
Db	30298	GAGGCTTGTATCCGAGCAATTTGGAAATTAAGGTAGAGAGAGATGACAGTACGCA	30239
QY	747	-----AATTACTATGACCAAAAGATGATGCTACA	779
Db	30238	TTGGAAAGCTAATAATATTGTATTCTATGCTGACACAAATTTGAGACTTGAAGCTGGA	30179
QY	780	ATTCAACTGGAAAAGGACTGCACTCTTTAAACACAGCGCTGCGTGTCTCTATAGTT	839
Db	30178	GTTTATTATTTGGCGCCGAAACCGCACTCTTCGATTAAGGAAGCTTTGGAAACTATTGTGCTCT	30119
QY	840	GTGTGTGAGAGATCCCATGGCAACTGTGGTAGAG-----GTCAGATCTAAGCCCAAGAG	893
Db	30118	GTGCGTGGCGCAACCCGATCTGATAGACGCTCGTGGAGAGCGTACACCGTTAAGCCCAACA	30055
QY	894	CAGTGTGGCGGCTCAAGCCTTTGGACACTGTGAGACTTGGAAAGCTGTGCTTCGAAAGTT	953
Db	30058	CAATAGTGGCGCCCAACTCCCTCGGATTCCTGTGGAAAGGGAAGCTTGCTGCGGGAACCT	29995
QY	954	GGAATTAATTTGCTAAAGGAACCATGAGATGTGCTGGAAGAGCTCTACTCAAGGGATCAAT	1013
Db	29998	CAGCACTATCTGCCAAGGAGACAGTGCATATAGCCGAAAATTTATATACCAGAAGTTTCAT	29933
QY	1014	CAGCTATCCCGCAACGAAACAATTTTTCOCAGAGACTTAAACCTGACGGTGTGGT	1073
Db	29938	CAGCTATCCCGTACGGAACCAACAGTTTCCAAAGGATTCGCTCGGACACACTAGT	29879
QY	1074	GGAACGACAGACCCCGATTCACAGCTGTGGGGGCTTTGGCCAGAGCAATCTAGAGCGGGG	1133
Db	29878	TGAGATGACAAAGGGGTACACGGGACGTGGGAGGATTTGGCCACAGCGGGTATTTGAGTGG--	29821
QY	1134	TGGTCTCACCCACCGCAAGCAATGGGGAACAAGTGTGACCAAGCTCAACCCCTCACCCGAC	1193
Db	29820	-GGACCAATTCCTCGAAACGCTTACAAATCGGATCAGGCCCATCTCCAAATTCATCCAC	29766
QY	1194	CAAA-----	1197
Db	29761	CAAGCTGGCAGAAAGTAAAGGAATCTTTATTTTAACTTAAACCGCAACATACATACCC	29702
QY	1198	--TACACCAACAACCTTACAGGGAGATGAACAGGAGCTGTACGAGTTTATTTGTCGCATTT	1255

Db	29701	TTTTCCCTAAAGACCTGGACGGGAAACGAGGCGCAGTGTACGAACCTGGTCTCCGGCACT	29642
QY	1256	TCCTGGCTTGGCTGCTCCAGGATGCTCAGAGGCGCAGSAGACACACTGTGAGATCGACATCG	1315
Db	29641	TCCTTCGCCCTGGGTCCAGCAAGATGTGAGTCGGCTCCGAAACCCCTTGCTATACGACATGG	29582
QY	1316	CTCAGGAACGGCTTTGTGGACCCATGGCCCTCATGATTTCTGGCCCGGAAACATATCGSAGTGGT	1375
Db	29581	CCGGGAGAAAGTTACACGGCCCAATGGTCTTGCTTTATTCACGACGAAGAACTACCTGATGTTT	29522
QY	1376	ATCCATATGATCACTGAGAGTGCACAGATCTCCCTGTCTATGAGAGCAAGATGCCACTTTC	1435
Db	29521	ATGTTTATGACAAAGTGGAGTGGCAGCAGATCCATCACTATGAAAAACGGGACGCTTTTG	29462
QY	1436	AGCCACGACACCGTGGAGATGTGGTGCAGCGGGAGACCAAGCCCAAGCTGCTACCCGAGG	1495
Db	29461	AGCCACACGAGATGTGCTGCCACGAAAGTGGCCACCACTGCCCCGCCCTCGTACCGAAG	29402
QY	1496	CCGACCTATTGGCCCTCATGTGAGAGAACATGGCATGTGTATGAGATGCCACTATGGGGAGC	1555
Db	29401	CGGATTTGATTGGCACTGTGATGGAAAAAGCATGGGATTGGTACAAATGCCACGACGCTGAGC	29342
QY	1556	ACATCGAGACCATCAAAACCCGAGATGATCGTGGGGCTCAACCCACAGACAAGGCTTCCTCC	1615
Db	29341	ACATTAACACCATCAACGAGACCGGATGCATATCGGAGTCTGTGATAGAGGCTTTTGGTAC	29282
QY	1616	CTGGGCAACCTGGGGCATGGGACTTGTGGAAGGTTATGATTCATGGGCTATGAAATGTCTA	1675
Db	29281	CTGGAGTATTGGATGGAGGACTGTATGAGGGGATACGATGCCATGGAGGTGGCCCTTGCA	29222
QY	1676	AGCCTGACCTCCGGGCTAATCTGCAAGCTGATCTTAAGCTCATCTGTGTATGSCAAAAAG	1735
Db	29221	AGCCACAACCTTCGTGCTGATGTTGGAATTTGAACTATTTGGCCAGGGGGCAGAAAG	29162
QY	1736	ACAATTTGTGTTCTAAGGCAGCAGCAAGTGCAGAAATACAGAGG	1780
Db	29161	ATCCCAAGGTGTCTGTGACCGACACGATAGCCAACTACAGAGG	29117

LOCUS	157766 bp	DNA	linear	INV 08-NOV-2003
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
AC009017	157766 bp	DNA	linear	INV 08-NOV-2003
Drosophila	melanogaster, chromosome 2L, region 37D-37E, BAC clone			
BACR27M12,	complete sequence.			
AC009017				
AC009017.1	GI:16798949			
fruit fly.				
Drosophila	melanogaster			
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
1 (bases 1 to 157766)				
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,				
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,				
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,				
Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,				
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D.,				
Fertieria,S., Frisze,E., Galle,R.F., Gard,N.S., George,R.A.,				
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,				
Idegawa,C., Jalali,M., Kruse,D., Li,P., Mettel,B., Moshrefi,A.,				
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunn,J.,				
Paclet,T., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,				
Phouneavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,				
Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M.,				
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.				
Sequencing of Drosophila chromosome 2L, region 37D-37E				
Unpublished				
2 (bases 1 to 157766)				
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,				
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,				
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,				
Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,				

TITLE	JOURNAL	COMMENT	FEATURES
Query Match	Best Local Similarity	Matches	
14.6%;	Score 433.4;	DB 3;	Length 157766;
55.7%;	Pred. No. 2.66-104;		
0;	Mismatches 671;	Indels 119;	Gaps 4;
BASE COUNT	44324 a	35420 c	35514 g 42508 t
ORIGIN			
112	CGAGAGAGAGACCTTTCAAAATTCACAAAGATCTATGATATTTGATTATCATCTGTATGGC	171	
85416	CAGGCGCAGAGCTACTCGTGTACACAAAGGCTTTCGATTTTCGAGGCCCTGTGGCCGT	85475	
172	CAGAGTGTACCATGGTATGATGACTTCATTTCTTGACATTTACAGCTCATGATTTCCAG	231	
85476	CAGATGCGCAAAATGGTTATATACCTCCGTATTCGCGCCACATGATGACAGCTGGATTTCCAG	85535	
232	ATGCGATTTCCAAATATGCGACAGCTGCACACCCCTCTTCTCTTTTGAAGCAGAAATTTGAA	291	
85536	GTGTCTTACAAAATTTGGAGAACGGTGTGATCCGGCTCTTTATTTCGACGCGCCGGTGAAA	85595	
292	AAGTACTGCCACAGAAATTTTGTAGACATCAAGAAAATTGGAAAGACAGACTCGCCAG	351	
85596	AAGGCGCTGGGCTGGGATTTAGAGCTATCAAAAAGAGACCTTGGAAACGCGAGTGGCGCGC	85655	
352	TGCGACGGTCTGGGATGTGACATGCAGCTGATATGAGAAAGCGAAACATGGGGTTTGA	411	
85656	TGTACGGGATGATTATCTGAGCGGATTCGATTCGCGGCGCGAGAAATTTGGCTACAG	85715	
412	ATTATTCACGTTGTATGAAGGCTGTATAAGCCCAATCTGCAGAGTGTTCGCGAACCCGATTCT	471	
85716	ATTATTTGATGTGTGTCGGGCTATCAAAACGAAATTTTGGTTATATCGTACCACTTTTCGG	85775	
472	GAGATCACACCCCATGCCGCTGAGACAGCTTTGTGAAAACCTGCACCGAGCCTGATTCAGG	531	
85776	GAGATTTACACAGCGTGGCGGTCTCTCGGCTCTTCGACAGCAATTTG886GACGCCGACAAAAG	85835	
532	GTGAGCCATGCTGTGGATGTTGAGGACAGAGCTGGACCGGAGGATTTGGAGCTGGCTTTCT	591	
85836	CAGAGCCATGAGTGTGATGTGCGACCGAATCGGATTTGTGCAACTGAGTGGCGCCATTCAC	85895	

TITLE	Direct Submission
JOURNAL	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
COMMENT	On Oct 9, 2000 this sequence version replaced gi:7298582.
FEATURES	location/Qualifiers
source	1..299449
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/chromosome="2L"
	join(619..833,3530..3863,3923..4604,4675..5176,5254..5430,5487..5895)
mrna	/gene="Sd"
	/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."
	/product="CT28175"
	/db_xref="FLYBASE:FBan0009999"
	/db_xref="FLYBASE:FBgn0003346"
	<619..>5895
gene	/gene="Sd"
	/note="CG9999"
	/map="37E2-37E3"
	/db_xref="FLYBASE:FBan0009999"
	/db_xref="FLYBASE:FBgn0003346"
	complement(join(1516..2125,2209..2429,2490..2628,2687..3005,3071..3339))
mrna	/gene="Hs2st"
	/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."
	/product="CT28767"
	/db_xref="FLYBASE:FBan0010234"
	/db_xref="FLYBASE:FBgn0024230"
	complement(<1516..>3339)
gene	/gene="Hs2st"
	/note="CG10234"
	/map="37E3-37E3"
	/db_xref="FLYBASE:FBan0010234"
	/db_xref="FLYBASE:FBgn0024230"
	complement(join(1882..2125,2209..2429,2490..2628,2687..3005,3071..3197))
CDS	/gene="Hs2st"
	/note="Hs2st gene product: Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"
	/codon_start=1
	/db_xref="FLYBASE:FBan0010234"
	/db_xref="FLYBASE:FBgn0024230"
	/protein_id="AAF53800.1"
	/db_xref="GI:7298583"
	/translation="MFRKLKMLLPRTMHLILALCAVTCAGYWLMSRIRLHAFKPLSKGDSLSPQHASSTTDFDEFEHLVLYLNRPVKTGSTSVNATYDLCKRNKTHVLRINVTANHHVLSLPNQIOFVRNVSRWHEMKPALYHGHMAFLDESKFOIAHKFIYINLVRRPLDLRLSYIYFLFEGDNYRPNLVRKAKNGKITDEECVOKOPDCDKNMQLQIPFCGHAACEWPEGSWALDQAKRLVNEVFLVGTQEOMYEFVLDLERSLRIFGFEFHYHNSNSHLRVTSSKLPPSESTIKSIQKTKIMQEMENDLDFALAQEPFKKKLQMQDNKHVOKPMYKIRPK"
	join(3764..3863,3923..4604,4675..5176,5254..5430,5487..5816)
CDS	/gene="Sd"
	/note="Sd gene product: Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript"
	/codon_start=1
	/db_xref="FLYBASE:FBan0009999"
	/db_xref="FLYBASE:FBgn0003346"
	/protein_id="AAF53801.1"
	/db_xref="GI:7298584"
	/translation="MSTFRASMAAOLGEOGISTFENKVLSTMTAADVDVVALNKOITVYILNDGNTLGEBAKATIGSLKHPEFRKALNMFTGRILSITPALKHLGALIVAGAKTIVLSDNALGNGMKGLELLRSPVCSYLOELLCLNCGLGEGSGMSLRALIDANANKAGFPLOLRFVIGSRNRLDAGATEMATLPLRLTREMSFGDCLIKTIYEGVEALAESEKRNPHLRVILNMNDNTLKSSEGAETIAEALPILRLREMSFGDCLIKTIYAHFGEALGSRNLEIYIDLGPNINDGILVYVNMGNPKLRIINLDGNSGEEBSFKTISMSKPLPTAALQIPQHOEEEDLEDEYQADKQADYEEEEEVHEHNDTTEAD
	EDSEGDDEDEDEDEEYSNAEETAVYTTNAVYTKLFENDTNSMASETPAVANKTISQKCPERKCLSQKPCSQEDFSLDMNKLQSLIVNDQDNLHLLVFTTLCAHLSSQSKALDLAVLSYQATFDYAIKQCTRYLANVLMQRLPLPKKEVHFSDYVKNCRFALREALKOPTEADINKNSKFTLEGAES"
	complement(join(6316..6639,6721..9758,9825..10002,10075..10248,10306..10863,11829..>12084))
mrna	/gene="Top2"
	/product="CT28703"
	/db_xref="FLYBASE:FBan0010223"
	/db_xref="FLYBASE:FBgn0003732"
	complement(<6316..>12084)
gene	/gene="Top2"
	/note="CG10223"
	/map="37E3-37E4"
	/db_xref="FLYBASE:FBan0010223"
	/db_xref="FLYBASE:FBgn0003732"
	complement(join(6316..6639,6721..9758,9825..10002,10075..10248,10306..10863))
CDS	/gene="Top2"
	/note="Top2 gene product"
	/codon_start=1
	/db_xref="FLYBASE:FBan0010223"
	/db_xref="FLYBASE:FBgn0003732"
	/protein_id="AAF53802.1"
	/db_xref="GI:7298585"
	/translation="MENGKALSTEDQYQKKSQLEHLLRPDSYIGSVETKELMNYDNSONRMVQKESISFVPGYLTFEDILVNAADNKKORDKSMNTIKIDIPERMVSVNNGOGIPVTMHKEQKMYPTMLIFGHLLTSDENYDDEKRYTGNGYGAKEKCNFTSFVYETATREYKSKSFQTMGNMNGKASDVQALITDNGTGYTRITFSPDLAKFMDLDDIYVALMSRRADVVASASKGVYFLNGKLTGIYHEVANAEMVYACCPDRGQOQVFNVSITATKSGRRHYDVYDNLKQLELYKNNKNNGKITKRPQVRNHHVYVYNKLLIEPTDSSOTKENNTLOOKFGSKCTLSEKFTINMSSGSIVESYLVAMAKFAQNDLITGGRKSSKIKGIPKLEDANEAGKNSIKCTLILFEGGSALVSGVIGRDLGVFPLKGLLNRVANFQKLOSENAENNNLCKIIGLOYKKYTLFJEDDCTRTYGVGMIMTDODDGSHTGLLINFITHNMPELRLPFLPEEITPILVATKKNEDLSYSLPEFEEMKNDPANHHTYNIKYVGLGTSKSEAKKEYFQDMDRHRLIFKYGSDVDSIYVAFSEKKNHSEKRVVLTNHMDPKRRKEKLEIPREYLYTKGTKITADPTNIEVLPFNADNRSRILPSYDGAKPCQKRVAFTEPKRNDKREYVAQULSSSVAEKSAVHHREVSLOMTIVYLAQNPVGNANINLEPRQGFSTRLSGSKDCASARYIFTIMSPLEFLVHPDLDDLDYQDDGQKTEPLNLLYSDGRYIOTGNIOILSGNRLSEISLVEGVYTOYKKNVLEPLSNGTEKYVGIISERYEHDTTVRYFVIFSPAGFEFRIHAEEGEYVFKLTTTLSTNOMHAREDONNCLRPFTALIDIKYVYKLRREYVAPRDFIVGLOALDRISDARFLTECKEKKIYENKORAKMCDLLKRGYRPDPYKQRIRIKMEDAQADEDEDEBEAAPSYSKAKKEVDPKAEKLTLYVKKFEDYLLGSMWMLTEKKNELLKQDITKUSELSLRKATPEMLVDDIDALBSKLINEVEKERAEEOGINLKTAKLKQKSSASAGRKYSMGGAGADVPDDGEPEYEFKITEEIIKKMAAAKVAQAAPKPKPKPKPKPKPKPKQIKAEPASGDVEDEFDAWEGSKTSPRAKKAIVYKKEPKKENGGLKOSKIDFSKAKAKSSDDVYEYTPRAPERGRQASKKIDYSSLFSPDEEDGNGVSDDGNASDDSPKRPAAKRGREDSESGKAKKAPPKRAVYESDDDDLEIDEDDDSDSDFNC"
	join(13443..13558,14987..15146,15208..15302,15402..15746,15806..15984,16054..>16171)
mrna	/gene="CG10026"
	/product="CT28189"
	/db_xref="FLYBASE:FBan0010026"
	/db_xref="FLYBASE:FBgn0032785"
	<13443..>16171
gene	/gene="CG10026"
	/map="37E3-37E4"
	/db_xref="FLYBASE:FBan0010026"
	/db_xref="FLYBASE:FBgn0032785"
	join((15011..15146,15208..15302,15402..15746,15806..15984,16054..16171)
CDS	/gene="CG10026"
	/note="CG10026 gene product"
	/codon_start=1
	/db_xref="FLYBASE:FBan0010026"
	/db_xref="FLYBASE:FBgn0032785"
	/protein_id="AAF53803.1"
	/db_xref="GI:7298586"
	/translation="MPTIEKLNITTEEVPEHILRLAQEOGECSSKDKYIEQFRNYILEHHQCPHNSDAKYLEKFLARARWKTLENSKLLCSYRRREQKSYEVKVRPLDLRH

[illegible]

Db	58055	TTGACATGCAAAAGGGGTCC	CCCCGACACTGGGGAGCAATTTTCC	ACGAGGGGATGTAGTGG--	58112
QY	1134	TGTCGTCGCCACCCACAGCAT	TGGGAACAGTCTGACCAAGCTC	ACCCTCCATTCACCCAC	1193
Db	58113	-GGAACAAATCTCTGAAACGG	TAAACAAATGGATCAGGCCAT	CTCTCCAAATTCATCCAC	58171
QY	1194	CAAA-----			1197
Db	58172	CAAGCTGGCAGAAAGTAAGG	AAATCTTTATATCTAAACCGCA	ACATTAACATACCCC	58231
QY	1198	--TACACCAACAACTTTAC	AGGAGATGAACACGCACCT	TACAGATTTATTTGGCCATT	1255
Db	58232	TTTCTCTTAAAGACTTCG	AGGGGAAACGAGGCGCAGAT	GTACAACTGGTCTCCGGCCT	58291
QY	1256	TCCTGGCTTTGGCTGCC	CAGATGCTCGGGGACAGAG	ACACAGTGAATGCACATCG	1315
Db	58292	TCCTCGCTCGGCTCAGCA	AAAGATGCATGCGGCTCC	GAAACCTTGTCCATATCCG	58351
QY	1316	CTCAGAAACGCTTGTGG	CCCCATGGCCTCATGATTT	CTGGCCGAAACTATCTGAT	1375
Db	58352	CCGGTAGAAGTTACGCG	CCCATGGTCTGGTTATC	ACGAAAGAACTACCTAGAT	58411
QY	1376	ATCCATATGATCACTGG	AGTGCACAAAGATCTCCT	GTGTATAGACAGATCCCACT	1435
Db	58412	ATGTTATGACAAAGGG	AGTGCACAAAGATCTAC	ATGATGAAAAACGGCAGCG	58471
QY	1436	AGCCAGCAGCAGCGTGA	TGGAGTGGGAGGAGAC	CGCCACCCAAAGCTGTCAC	1495
Db	58472	AGCCACGGAAGTGTGCT	CCACGAAAGTGTGCAC	CACTGCCCCCTGCTGAC	58531
QY	1496	CCGACCTCATTTGCCCT	CATGTGAGAAAGCAT	TGGCATTTGGTATGCAC	1555
Db	58532	CGGATTTGATTGCACT	GTATGGAAGAAAGCAT	TGGGATTTGGTATGAC	58591
QY	1556	ACATCGAAGACCATCAA	AGCCGGATGTACGTGGCC	CTCACCACAGACGGTTCT	1615
Db	58592	ACATTAACACATCATCA	AGGAAGCGGATCATCG	GGTCTGGATTAAGGCTTT	58651
QY	1616	CTGGCACCCTGGGGCAT	TGGGACTTTGGAAGGTT	ATGATTTTCATGGCCTA	1675
Db	58652	CTGGAATTTATTGGA	TGGAGCTGTATGTGGAG	ATACGATGAGCTGGCC	58711
QY	1676	AGCCTGACCTCCGGGCT	GAACTGAAAGCTGTAA	AGCTGATCTGTGATGGCA	1735
Db	58712	AGCCACAACTTCGTG	CTGATGTTTGAATCT	GTGAACCTAATTTGGCA	58771
QY	1736	ACAAATTTGTGCTTAA	AGCAGCAGCATGTCA	GAAGTAACAGCAGG	1780
Db	58772	ATCCCAAGGTGTCTC	TGACCGAACAGATAG	CAACAGTACMACCAGG	58816

RESULT 11

AC018257 73907 bp DNA linear HTG 09-DEC-1999

LOCUS AC018257 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered

DEFINITION AC018257 pieces.

ACCESSION AC018257

VERSION AC018257.1 GI:6552934

KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 73907)

AUTHORS Adams,M. and Venter,J.C.

TITLE Direct Submission

JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM.10214069 by the submitter. For more information on this record e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

FEATURES	Location/Qualifiers
source	1. .73907

BASE COUNT	20424	a	17157	c	17162	g	19164	t
ORIGIN								

Query Match	14.4%	Score 429.4	DB 2	Length 73907
Best Local Similarity	56.0%	Pred. NO. 2.7e-103		
Matches 999	Conservative	0	Mismatches 666	Indels 120; Gaps 5

QY	112	CGGAGGAAAGACATTTCAAAATTCACAAGACTCTGTGATTTGATTAATCACTGATATGC	171
Db	5454	CAGGGGAAAGCTACTCGGTGTACACAAAGGTCTTGATTCGAGGCCCTGTGGGGGT	5513
QY	172	CAGATGTTTACCATGTTATGACTTCTAGTTTCTGCAATTTACTGGTCTATGATTTCCAG	231
Db	5514	CAGATATCCAAAATGGTTATGACCTCCGATCCGGCCACATATCAGACTGGCAATTCAG	5573
QY	232	ATGCAGTTTGGAAAATGAGCAGAGCTCAACCCCTGTGCTCTTGTGAAGCAGAAATTCAA	291
Db	5574	GTC-TCTTAAAAAATTTGAGAAACGGTGGATCCGGCTCTTATATGAGCCGCCGTGA	5632
QY	292	AAGTACTGCCAGAGATTTTGTAGACATCAAGAAAACTTTGGAAACGAGACTGGCAG	351
Db	5633	AAGGGCTGGGCTCGGATTTGTAGAGCCATCAAAAAGAGCTTGGAAAGCGAGGTGGCGGC	5692
QY	352	TGCGAGCTGTGGTGATCTTGACATGACTGTCGTGATAGGAAGGCGAAACATCGGGTTTAC	411
Db	5693	TGTGAGGATTTGATTTATCTGTGACGGATTTGCGATCGGAGGGCGAACAATTTGCTACGAG	5752
QY	412	ATTATCCAGCTGTGTAAAGCTGTAAAGCCCAATCTGCAGGTGTTCCGAGCCGATTCCT	471
Db	5753	ATTATGATGTGTGCGCTATCAACCGAAATTTTGGTTATCTGTGCCACTTTCG	5812
QY	472	GAGATCACACCCATGCCCTGACGAGACAGCTTGTGAAAACCTGACCGACCTGTACAGG	531
Db	5813	GAGTTTACACAGGTGGGGGCTCGGGCTGTGCAGCAATTTGGGACCCGGACAAAAG	5872
QY	532	GTCGAGCATCTGTGATGTGAGGACGAGAGCTGAGACTGAGATTTGAGCTGCTTACT	591
Db	5873	CAGAGCGATGATGATGTGTGCGCCACAGAACTGTGATTTCCAACTGTGTCCGCATCAC	5932
QY	592	AGGTTCCAGACCTCGCGCTTCAGAGATTTTCTGAAGTGTGGCAGAGACGCTATC	651
Db	5933	CGGTTTCAGACAATGCGGCTACAACTGTTCCCGGAAAATATGAGTAAAGCTCTA	5992
QY	652	AGTTAGGCGACGTCCAGTTCCCAACAGGGCTTTGTGTGACAGCGGTTGAAAGCATT	711
Db	5993	TCCATGGAAGCTGCGCAAAATTCACACCCCTGGGATTTGTGGCCGAAGGTTACAGAGATT	6052
QY	712	CAGGCTTTTGTACCAAAATCTTCCACAAATAA-----	746
Db	6053	GAGGCGTGTATCGAGCGATTTTGAAGATTAAGAGGTAAGAGAGGATGCAGTAGSCA	6112
QY	747	-----AGTAATCATGACCAACAAGATGTATCTAGA	779
Db	6113	TTGGAAGCTATTAATATTGTATTCTATTCTGACACAATGTATGACTTGAGCGTGA	6172
QY	780	ATTCAACTGGAAGAAGCATATGACTCTTTTAAACACAGCGCTTGCTCTCTATCAATGTT	839
Db	6173	GTTTAATTTGGGCCCGGAACCGACTCTTTCATTAAGGAAGCTTGTGAACAATTTTGTCTCT	6232
QY	840	GTCGTGTGAGAGATCCCATGGCAACTGTGTAGAG-----	893
Db	6233	CTGCCTGTGGCAACCGATCTAGAGCGCTCTGTGAAGAGCGTCAACCGTTAAGCCCAACA	6292
QY	894	CAGTGGCGGCTCAAGCCTTGGACACTGTGAGACTTTAGAAGCTGTGCTTCTCGAAATTT	953
Db	6293	CAATATGGCGCCCAACTCGCTGGATTAACGTGGAAATGGAAGAGCTTTGGCTGTGAGAAACT	6352

QY	954	GAGATTAATGTTGTAAGAAGAACCATAGGATTGCTGAGAAAGCTGTACACTCAAGGTCAT	1013
Db	6353	CNAGCTATCTGTGCGAAGGAGACGATGACATATGACCGAATAATTATATACCAAGGGTTTCAT	6412
QY	1014	CAGCTATCCCGGAACGAGAACAACATTTTTTCCAGAGACTTAAACCTGAGCGTGTGGT	1073
Db	6413	CAGTATATCCCGTACGGAACCAACCACTTCTCCAAAGAGTTCCGCTGCGCAACCACTAGT	6472
QY	1074	GGAAACGACAGACCCCGCATCCACGCTGGGGGGGCGCTTTGGCCAGAGCATTTCTAGAGCGGG	1133
Db	6473	TGAGATGCGAAGACGGGTCTACCCGACTGGGGAGCAATTTGGCCAGGGGGGATTTAGTGG--	6530
QY	1134	TGTGTCCACCCACGACATGGGAACAACTGTAGACAAACCTCAACCTCCATTCACCCAC	1193
Db	6531	GGACCAATTCCTCGAAACGGTAACAAATTCGAGATCAGGCCCATCTCTCCAAATTCATCCAC	6589
QY	1194	CMAA-----	1197
Db	6590	CAAGCTGGCAGAAAGTAAGGAAATCTTTATCTTAAATCCGAACATATACATACCCC	6649
QY	1198	--TACCAACACACTTACAGGGAGATGAAACGCGACTGTACAGATTATATGTTGGCATT	1255
Db	6650	TTTCTCTTAAGACTGTGAGGAAACGAGCGCGAGTGTACAACTGGTCTCCGGCACT	6709
QY	1256	TCTGCGCTTGTGCTCCACGAGATGCTCAGGGGACGAGAACCAAGTGGAGATCGACATCG	1315
Db	6710	TCTCTGCGCTGGGTGACGAAAGATGAGTCGGCTCCGAACCTTGTCCATATCGACATTTG	6769
QY	1316	CTCAGGAACCGCTTTGTGGCCCATGGCTCATGATTTCTGCGCCGAACATATGTGATGTGT	1375
Db	6770	CCGGTGAACATTCACGGCCCAATGGTCTGGTTATTCAGGAACCACTTACCTAGATGTTT	6829
QY	1376	ATCCATATGATCAGTGGAGTGTACAAAGATCTCCCTGTCTATGACAAAGATCCCATTTTC	1435
Db	6830	ATGTTTATGACAAAGTGGAGTGCACAGACATTCATCATTATGAAGAACGGCGAGGTTTG	6889
QY	1436	AGCCAGCAGCCGTGGAGATGTTGAGCGGGGAGACCAAGCCCAAGCTGCTCACCGAGG	1495
Db	6890	AGCCACGAGAGTGTGCTCTCCAGAAAGTGTCCACCACTGCCCCCTGCTGACCGAAG	6949
QY	1496	CCGACCTCATTTGCCCTCATGAGAAAGCATGGCATTTGGTATGCGCATCTACGGGAGC	1555
Db	6950	CGGATTTGATTGACATGATGGAAAGCATGGATTGGTATGACGATGCCACGACGCTGAGC	7009
QY	1556	ACATCGAGACCATCAAAAGCCCGGATGTACGTTGAGGGGCTCACCCCGACGAACAGGGTTCTCC	1615
Db	7010	ACATTAACCATCATTAAGGAGACGCGGATATCATCGAGTGTCTGATTAAGGCGCTTTTGGTAC	7069
QY	1616	CTGGGCACTTGGGCGCATGGGACTTGTGAAAGGTTTATGATTTCCATGGGCTATGAATGTCTA	1675
Db	7070	CTGGAGTATTGGAATGGGACTGTATGAGGGATATGATGCCATGAGACTGGCCCTTCCCA	7129
QY	1676	AGCCTGACCTCCGGGCTGTACTGTGAAGCTATCTGAAAGCTGATCTGTGATGTGCAAAAGG	1735
Db	7130	AGCCACAACCTTCGTGTGATTTGAATTTGGAATTCGAAACTATTTTGCAGGGGCGAAGG	7189
QY	1736	ACAATTTGTTGGTTCTAAGCGAGCAAGTCCGAATATCAAGCAGG	1780
Db	7190	ATCCCAAGGTGTCTCTGACCGACAGATAGCCAAAGTACAAAGCAGG	7234
RESULT 12			
AF057032			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditiida;			
AF057032			
2422 bp			
mrna			
linear			
INV 27-Apr-2000			
Caenorhabditis elegans DNA topoisomerase III (TOP3) mRNA, complete			
cgs.			
AF057032			
AF057032.1			
GI:3047012			
Caenorhabditis elegans.			
Caenorhabditis elegans			
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditiida;			

[illegible][illegible]

QY	9111	CTTGGACACTGTGGAGCTTGAAGAAGCGGCTTCGCAAAAGTGGAAATTAATGCTAAAG	970
Db	1012	CCCTGAAACACTGTGGAGATGCTGCGGTGTGGCCAGCTCTTCTCTGGCATGTGGGCCCGCAGC	1071
QY	971	AAACCATGAGGATGCTGAGAAAGCTTACACTCAAGGGTACATCAGTATCCCGAACAG	1030
Db	1072	ACGGCATGCACAGCGGTGAGGGGCTCTACACGCAAGGCTACATGAGCTATACCCACGAGCAG	1131
QY	1031	AAACAAACATTTTCCAGACACTTAAACCTGCAGCGGTGTGGTGGAAACAGCAAGACCCCG	1090
Db	1132	AGACACCCACTACCTGTGAGAACTTGTAGCTGAAGGGCTCTGTGGGAGGAGGCAAGCAAC	1191
QY	1091	ATCCACGCTGGGGGCGCTTTTCCAGACACTTCTAGACCGGGGTGGTCCACCCACGACGA	1150
Db	1192	ACCCTACTGGGCCGACACGGGTGAAGCGGTGTTAGCAGAAGGTATCAACCGCCCGCGGA	1251
QY	1151	ATGGGAACAAGTCTGACCAAGCTCAACCTCCCATTCACCCACCAATATACACCA --CA	1207
Db	1252	AAGGCATGAGCGCGCGCACCATCCCCCATACCCCCATGAAGTCTCCACAGAGCGG	1311
QY	1208	ACTTACAGGAGAGTAAACAGGAGACTGTACGAGTTATGTGTGCGCATTTCCGTGGTGTCT	1267
Db	1312	AATTGAGGGGTGACCGGTGGGCGCTCTATGTAGTATCATACACAGACTTCTACGACAGG	1371
QY	1268	GCTCCAGAGTGTCTCAGGGGAGAGACACAGTGGAGATGCACATCTGCTCAGAACGCT	1327
Db	1372	TCAGCACTGAGTGTCAAGTACCTGCAGAGACACCATCTCTTGAGATTTGGGCCGAGCTCT	1431
QY	1338	TTGTGGCCCATGTGGCTCTGATATCTGGGCCCAAACTATCTGATGTGATCATATGATTC	1387
Db	1432	TCACCTGTCTCGGGAGAACCGCTCTCTACCAAGGTTCTACGGAAGTCAATGCCCTTGGCAGA	1491
QY	1388	ACTGGAGT --GACAAAGATCTCTCCCTGTCTATGAGCAAGATATCCACTTTGACGCCACGA	1444
Db	1492	GCGTGCCCTGGAGAGAGAGCTGCCCACTGTGCACGCGGGGTGATGCTCTCCCTGTGGGG	1551
QY	1445	CCGTGGAAATGTGTGACCGGGAGACACACCCAAAGCTGTCTACACGAGGCCACCTCA	1504
Db	1552	AGGTGAAGATCTGTGAGAGAGACAGAAACCAACCCGATACCTGTGACGAGGCGGAGCTCA	1611
QY	1505	TTGCCCTATGAGAGACATGGCATTTGGTATGGATGATCCACTATCAGGACATGACGAGA	1564
Db	1612	TCACGCTATGAGAGAGATGTGCATCGGACAGGATGCCAGCATCCCTGTGATATCAACA	1671
QY	1565	CCATCAAAAGCCGGATGTACGTGGGCTCAACCCAGACAGCGGTTCTCCCTGGGGACC	1624
Db	1672	ACATGTGCAGAGGCAACTATGTCTACAGGTGAGAGAGCGGGCGCGCTCAAGCCCAACCAAC	1731
QY	1625	TGGGAGTGGAGACTTGTGAAGGTTATGTATTCATGGGCTATGAATAATGTCTAAGCCTGAC	1684
Db	1732	TCGGCATGTCTGTGTGACCGGCTACTATATAGATTGATGCAAGCTGTGTCTCCCAACA	1791
QY	1685	TCGGGCTGAACCTGGAAGCTGATCGAAGCTGATGTGTATGGGCAAAAAAGCAAAATTTG	1744
Db	1792	TCGCGAGTGAAGTGAAGAAGAGCTGAACCTGATGCGCCAGGAGCAAGGCGCCTACCGC	1851
QY	1745	TGTGTTCAAGCAGCAAGTGCAGAAATACMAAGCTTTTC	1785
Db	1852	AGGTCTCTGGGCGACACCTGTGACGTTCTCAAGAGGAAGTCT	1892

[illegible]

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE
1 (bases 1 to 2793)	Mammalia: Primates; Catarrhini; Homnidae; Homo.				
Ng,S.W., Liu,Y., Hasselblatt,K.T., Mok,S.C. and Berkowitz,R.S.		A new human topoisomerase III that interacts with SSI protein			
99128286		Nucleic Acids Res. 27 (4), 993-1000 (1999)			
2 (bases 1 to 2793)	Ng,S.W.	Direct Submission			
Submitted (09-MAR-1998)	OB/GYN & Reproductive Biology, Brigham & Women's Hospital, 221 Longwood Avenue, Boston, MA 02115, USA				
Location/Qualifiers					
1..2793					

[illegible]

```

Db 475 AGGGGAGAACATCTGCTTTGAGGTTCTTGATGCTGTTCTGCCCCGTCATGTAACAGAGCCC 534
449 AG-----GTTTGGAGCCCGATTTCTCGTAGATCAACCCCATCGCTCA 493
535 ATGTGGCGAAGAACCGTGTTCGGGGCAGGTTTACCTCATACGACACAGACATCT 594
494 GGACAGCTTGTAAACCTGACCGCCTGATCAGAGGGTGAAGGATGCTTGTGATGTA 553
595 GTATGCGCATGGCCTGCTAGGCGCCTGACCAACGAGGCGCTCTCATGTGATGCTC 654
554 GGACAGAGTGGACCTGAGATTTGAGAGCTGCTTACTAGTTTCCAGACCTGGGGCTC 613
655 GCGAGGAGCTGGACCTCGAATCGCTGTGATTCACAGAGTTTCAGATTAATATTTCC 714
614 AGAGATTTTTCCTGAGGCTGCGACAGAGCTCATCAGTTAGGAGAGCTGCGAGTTCC 673
715 AGGGGAATACGTGA---TTTAGACAGCTCTCTCATCTCTTTGGCGCGGTGACAGCTC 771
674 CCACACTGGCCTTGTGTGGAGCGGTTCAAGGCCATTCAGGCTTTTGTACCAAAATCT 733
772 CAACCCCTGGGATTTCTGTGTGAGAGACATGATTAATAATCCATCCTTCAACAGAGACT 831
734 TCCACAGATTTAAAGTACATCATGAC---CACAAAGATGGTATGTAATTCAACTGGA 790
832 ACTGGGTGCTGACAGGCCAAGGTTTACACTGACAAAGACAGATCTCTCTTTGGAGCTGG 891
791 AAAGCATGCACTCTTTTAAACACAGGCTTGCCTTATCTTATCAGTTGTGTGGAGG 850
892 ACCGATGAGAGTGTGTGACCGGAGATCGACAGATGTTTAAACATGACAAAGCTGG 951
851 ATCCCATGCACTGTGTGAGGTGACATCTAAGCCCAAGAGCAATGGCGGCTCAAG 910
952 AGAAGAAAGCCAGGTGAGAGCCCAAGAAAGAAAGAAAGCAACAGAGGCGCCCTGG 1011
911 CCTTGACACTGTGAGCTTGTAGAAAGCTGGCTTCTGCAAAATTTAGATTAATGCTAAG 970
1012 CCTGAAACACTGTGTGATATGCTGTGTGTGCGCACTCTTCTGTGGCATGGGGCGCAGC 1071
971 AAACATGAGGATTTGTGAGAAAGCTTACACTCAAGGATCAGTATGCCGGAAC 1030
1072 ACGCATGACAGAGCGCTGAGGGGCTTACACGAAAGCTCATCAGCTACCCAGCGAG 1131
1031 AAACAAACATTTTCCAGAGACTTAAACTGACGGTGTGTGTGGAACAGACACCCCG 1090
1132 AGACACACCTACTACCTGAGAACTTGTGAGAGGCTCTGTGCGCAGAGAGCCCAAC 1191
1091 ATCCACGCTGGGGGCGCTTGTGCCAGAGCAATTCAGAGCGGGTGTGCCACCCAGCA 1150
1192 ACCCTTACTGGCGGACACGCTGAGCGGTTGTTAGCAAGAGTATCAACCGCGCGGA 1251
1151 ATGGGAACAAGTGTGACCAAGCTCACCTCCATTCACCCACCAATAACACAA---CA 1207
1252 AAGGCCATGACGCGCGGACCATCCCCCAACCCCATGAAATGTGCCACAGAGCGG 1311
1208 ACTTACAGGAGATGAGAACAGCACTGACAGTATGTTTGTGCCATTTCTGGCTTGT 1267
1312 AATTAGGGGGTGAAGCGCTGTGCGGCTTATAGTACATCACAGACCTTATTCGCCACG 1371
1268 GCTCCCGAGATGCTCAGGGGACAGAGACACAGTGGAGATGACATGCTCAGAGAGCT 1327
1372 TCAGCCATGACTGCAATCTGACAGAGACACATCTCTTACAGAAATGGCGCGAGCTCT 1431
1328 TTGTGGCCCATGGCCTCATGATTTGCGCCGAAATATCTGGATGTGTATCATATGATC 1387
1432 TCACCTGTCTCCGGGAAGACGCTCTTACACAGGCTTACAGAGAGTATGCTCGGAGA 1491
1388 ACTGAGT---GACAGATCTCTCTGTCTATGAGCAAGATCCCACTTTCAGCCGCA 1444
1492 GCGTGGCCCTTGAGGAGAGAGCTGCGCACTTGCAGCGGGTGTATGCTTCTGTGGGG 1551
1445 CCGTGGAGATGTGAGAGGAGAGACAGCCACCAAGCTGCTCACCAGGAGCGCACTCA 1504

```

```

Db 1552 AGTGAAGATCTGTGAGAGACAGACGAACCAACCCGACTGACGAGAGCGGAGCTCA 1611
1505 TTGGCCTCATGAGAGAGATGGCATTTGGTACGATGCTCATCATCGAGACATCGAGA 1564
1612 TCAGCTCATGAGAGAGATGGCATGGCAGCGAGATGCCAGCATCCCTGTGATATCAACA 1671
1555 CCATCAAAAGCCGGATGTAGCTGGGCTCACCCCAAGACAGAGGTTCTCTCCCTGGGACC 1624
1672 ACATCTGCCACCGCAACTATGTACAGGTGAGAGAGCGGCGGCTCAAGCCCAACCAAC 1731
1625 TGGGCATGGAGCTTGTGAGAGATGATTTCCATGAGGCTATGAATGTCTAAGCTGACC 1684
1732 TGGCATGCTCTGTGTGACGCTTACTTAAAGATTGATGACAGAGCTGCTCCACCA 1791
1685 TCCGGGCTGAACGTGGAAGCTGATCTGAAGCTGATGATGCAAAAAAGACAAATTTG 1744
1792 TCCGAGATGCAAGTGAAGAGAGCTGAACCTGATGATGCCACGAGCGGCGACGCTACGCC 1851
1745 TTGTTTAAAGCAGCAGTGCAGAAATACAAAGAGTTTC 1785
1852 AGGTCTTGGGCGCACACCCTGGACGTTTCAAGAGAGATTG 1892

RESULT 15
BC002432 3133 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, topoisomerase (DNA) III beta, clone MGC:1867
DEFINITION IMAGE:3346377, mRNA, complete cds.
ACCESSION BC002432
VERSION BC002432.1 GI:12803238
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3133)
REFERENCE 1
AUTHORS Strausberg,R
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT Contact: MGC help desk
Email: gcaps-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: http://www.nisc.nih.gov/
nisc.mgc@hgrl.nih.gov
Contact: Shevchenko,Y., Wetherby,R.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
Lim,M., Maduro,Q.L., Mastello,C., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stanlipop,S., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tsurgon,C., Vogt,J.L., Walker,M.A.,
Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 5 Row: e Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4102878.
location/Qualifiers
1. 3133
/organism="Homo sapiens"
/db_xref="LocusID:8940"
/db_xref="taxon:9606"
/clone="MGC:1867 IMAGE:3346377"
/tissue_type="Skin, melanotic melanoma."
FEATURES
source

```


